



**Europäisches  
Patentamt**

**European  
Patent Office**

**Office européen  
des brevets**

**Bescheinigung**

**Certificate**

**Attestation**

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

**Patentanmeldung Nr. Patent application No. Demande de brevet n°**

03078772.5

**BEST AVAILABLE COPY**

Der Präsident des Europäischen Patentamts;  
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets  
p.o.

**R C van Dijk**



Anmeldung Nr:  
Application no.: 03078772.5  
Demande no:

Anmeldetag:  
Date of filing: 01.12.03  
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

ViroNovative B.V.  
Burgemeester Oudlaan 50  
3062 PA Rotterdam  
PAYS-BAS

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:  
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.  
If no title is shown please refer to the description.  
Si aucun titre n'est indiqué se référer à la description.)

Novel atypical pneumonia-causing virus

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s)  
revendiquée(s)

Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

EP/18.11.03/EP 03078613

Internationale Patentklassifikation/International Patent Classification/  
Classification internationale des brevets:

G12N7/00

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of  
filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL  
PT RO SE SI SK TR LI

01 DEC 2003

Title: Novel atypical pneumonia-causing virus

112

5

The invention relates to the field of virology.

Recently, a respiratory illness (atypical pneumonia) was diagnosed in an 8 months old patient that could not be attributed to SARS (Severe Acute Respiratory Syndrome) virus or any other known viral infection. The patient tested negative for influenza, parainfluenza, mumps and RSV and yet the disease was identified to be caused by a virus which closely resembled SARS.

For being able to trace its origin, monitor its epidemiology and prevent possible spreading of the disease, it is of great importance to be able to recognise viral causes of pneumonia in an early stage. Especially, if severe diseases are found to be caused by viruses, it is necessary to detect the identity of the virus as soon as possible, in order to develop diagnostic tools and possibly therapies. The SARS epidemic has shown that it is paramount for prevention of spread of the disease to be able to get an early diagnosis in order to timely take effective isolation measures and initiate quarantine precautions. Only then, world-wide contaminations can be prevented.

Furthermore, identification of the viral cause for the disease enables development of vaccines, which can be used prophylactically to protect people who are at risk of being infected. And, finally, knowledge of the viral cause enables to develop therapeutic measures.

Thus, there is great need in developing diagnostic tools and therapies for viral pneumonias in general, and particular to a novel disease-causing infectious agent, especially when this agent appears to be a virus.

The invention provides the nucleotide sequence of an isolated essentially mammalian positive-sense single stranded RNA virus belonging to the Coronaviruses, which is the causative factor for the new disease, hereinafter referred to as EMCRCoV and the disease being referred to as EMCRCoV-caused pneumonia. From a phylogenetic analysis of the Matrix and Nucleocapsid gene sequences of the virus (Fig. 2a and 2b) it appears that the virus is a distinct member of the group formed by PEDV (porcine epidemic diarrhea virus), HCoV-229E (human coronavirus 229E), PRCoV (porcine

respiratory coronavirus), TGEV (transmissible gastroenteritis virus), CaCoV (Canine coronavirus) and FeCoV (feline coronavirus). Based on amino acid identity matrices, human coronavirus 229E seems to be the closest relative (for all ORFs with the exception of Matrix which appears to be slightly more closely related to PEDV – see Figure 3).

Although phylogenetic analyses provide a convenient method of identifying a virus, several other possibly more straightforward albeit somewhat more coarse methods for identifying said virus or viral proteins or nucleic acids from said virus are herein also provided. As a rule of thumb an EMCR-Coronavirus can be identified by the percentages of homology of the virus, proteins or nucleic acids to be identified in comparison with viral proteins or nucleic acids identified herein by sequence. It is generally known that virus species, especially RNA virus species, often constitute a quasi species wherein a cluster of said viruses displays heterogeneity among its members. Thus it is expected that each isolate may have a somewhat different percentage relationship with the sequences of the isolate as provided herein.

When one wishes to compare a virus isolate with the sequences as listed in figure 1molo, the invention provides an isolated essentially mammalian positive-sense single stranded RNA virus (EMCR-CoV) belonging to the Coronaviruses and identifiable as phylogenetically corresponding thereto by determining a nucleic acid sequence of said virus and determining that said nucleic acid sequence has a percentage nucleic acid identity to the sequences as listed higher than the percentages identified herein for the nucleic acids as identified herein below in comparison with PEDV, 229E, PRCoV, TGEV, CaCoV and FeCoV. Likewise, an isolated essentially mammalian positive-sense single stranded RNA virus (EMCR-CoV) belonging to the Coronaviruses and identifiable as phylogenetically corresponding thereto by determining an amino acid sequence of said virus and determining that said amino acid sequence has a percentage amino acid homology to the sequences as listed which is essentially higher than the percentages provided herein in comparison with PEDV, 229E, PRCoV, TGEV, CaCoV and FeCoV.

With the provision of the sequence information of this EMCR-Coronavirus (EMCR-CoV), the invention provides diagnostic means and methods, prophylactic means and methods and therapeutic means and methods to be employed in the diagnosis, prevention and/or treatment of disease, in particular of respiratory disease (atypical pneumonia), in particular of mammals, more in particular in humans associated with infection by this virus. In virology, it is most advisory that diagnosis, prophylaxis and/



treatment of a specific viral infection is performed with reagents that are most specific for said specific virus causing said infection. In this case this means that it is preferred that said diagnosis, prophylaxis and/or treatment of an EMCR-CoV virus infection is performed with reagents that are most specific for EMCR-CoV virus. This by no means  
 5 however excludes the possibility that less specific, but sufficiently cross-reactive reagents are used instead, for example because they are more easily available and sufficiently address the task at hand.

The invention for example provides a method for virologically diagnosing an EMCR-CoV infection of an animal, in particular of a mammal, more in particular of a  
 10 human being, comprising determining in a sample of said animal the presence of a virus isolate or component thereof by reacting said sample with an EMCR-CoV specific nucleic acid or antibody according to the invention, and a method for serologically diagnosing EMCR-CoV infection of a mammal comprising determining in a sample of said mammal the presence of an antibody specifically directed against an EMCR-CoV virus or  
 15 component thereof by reacting said sample with an EMCR-CoV virus-specific proteinaceous molecule or fragment thereof or an antigen according to the invention.

The invention also provides a diagnostic kit for diagnosing an EMCR-CoV infection comprising an EMCR-CoV virus, an EMCR-CoV virus-specific nucleic acid, proteinaceous molecule or fragment thereof, antigen and/or an antibody according to the  
 20 invention, and preferably a means for detecting said EMCR-CoV virus, EMCR-CoV virus-specific nucleic acid, proteinaceous molecule or fragment thereof, antigen and/or an antibody, said means for example comprising an excitable group such as a fluorophore or enzymatic detection system used in the art (examples of suitable diagnostic kit format comprise IF, ELISA, neutralization assay, RT-PCR assay). To  
 25 determine whether an as yet unidentified virus component or synthetic analogue thereof such as nucleic acid, proteinaceous molecule or fragment thereof can be identified as EMCR-CoV-virus-specific, it suffices to analyse the nucleic acid or amino acid sequence of said component, for example for a stretch of said nucleic acid or amino acid, preferably of at least 10, more preferably at least 25, more preferably at least 40  
 30 nucleotides or amino acids (respectively), by sequence homology comparison with the provided EMCR-CoV viral sequences and with known non-EMCR-CoV viral sequences (human coronavirus 229E is preferably used) using for example phylogenetic analyses as provided herein. Depending on the degree of relationship with said EMCR-CoV or non-EMCR-CoV viral sequences, the component or synthetic analogue can be identified.

The invention thus provides the nucleotide sequence of a novel etiological agent, an isolated essentially mammalian positive-sense single stranded RNA virus (herein also called EMCR-CoV virus) belonging to the Coronaviridae family, and EMCR-CoV virus-specific components or synthetic analogues thereof.

5           Coronaviruses were first isolated from chickens in 1937, while the first human coronavirus was propagated *in vitro* by Tyrell and Bonoe in 1965. There are now about 13 species in this family, which infect cattle, pigs, rodents, cats, dogs, birds and man. Coronavirus particles are irregularly shaped, about 60-220 nm in diameter, with an outer envelope bearing distinctive, 'club-shaped' peplomers ( about 20 nm long and 10  
10 nm wide at the distal end). This 'crown-like' appearance give the family its name. The envelope carries two glycoproteins: S, the spike glycoprotein which is involved in cell fusion and is a major antigen, and M, the membrane glycoprotein, which is involved in budding and envelope formation. The genome is associated with a basic phosphoprotein, designated N. The genome of coronaviruses, a single stranded positive-sense RNA  
15 strand, is typically 27-31 Kb long and contains a 5' methylated cap and a 3' poly-A tail, by which it can directly function as an mRNA in the infected cell. Initially the 5' ORF 1 (about 20 Kb) is translated to produce a viral polymerase, which then produces a full length negative sense strand. This is used as a template to produce mRNA as a 'nested set' of transcripts, all with identical 5' non-translated leader sequence of 72 nucleotides  
20 and coincident 3' polyadenylated ends. Each mRNA thus produced is monocistronic, the genes at the 5' end being translated from the longest mRNA and so on. These unusual cytoplasmic structures are produced not by splicing, but by the polymerase during transcription. Between each of the genes there is a repeated intergenic sequence – AACUAAAC – which interacts with the transcriptase plus cellular factors to splice the  
25 leader sequence onto the start of each ORF. In some coronaviruses there are about 8 ORFs, coding for the proteins mentioned above, but also for a heamagglutinin esterase (HE), and several other non-structural proteins.

Newly isolated viruses are phylogenetically corresponding to and thus taxonomically corresponding to EMCR-CoV virus when comprising a gene order and/or  
30 amino acid sequence and/or nucleotide sequence sufficiently similar to our prototypic EMCR-CoV virus. The highest amino acid sequence identity, between ORFs of EMCR-CoV virus and any of the known other viruses of the same family to date are with human coronavirus 229E or Porcine Epidemic Diarrhea Virus (see Figures 3 and 4). The amino acid identities with human coronavirus 229E ranges from 45% (Nucleoprotein) to 81%

(Replicase 1b); interestingly, Replicase 1a has an identity of just 56% contrasting with Replicase 1b's 81% identity. EMCR CoV has a closer identity with human coronavirus 229E than with any of the known other viruses of the same family to date for all putative ORFs, with the exception of Matrix, which is slightly more closely related to the Matrix ORF of PEDV. Individual proteins or whole virus isolates with, respectively higher homology than these mentioned maximum values are considered phylogenetically corresponding and thus taxonomically corresponding to EMCR-CoV virus, and generally will be encoded by a nucleic acid sequence structurally corresponding with a sequence as shown in figure 1. Herewith the invention provides virus phylogenetically corresponding to the isolated virus of which the sequences are depicted in figure 1.

It should be noted that, similar to other viruses, a certain degree of variation can be expected to be found between EMCR-CoV-viruses isolated from different sources.

Also, the viral sequence of the EMCR-CoV virus or an isolated EMCR-CoV virus gene as provided herein for example shows less than 95%, preferably less than 90%, more preferably less than 80%, more preferably less than 70% and most preferably less than 65% nucleotide sequence homology or less than 95%, preferably less than 90%, more preferably less than 80%, more preferably less than 70% and most preferably less than 65% amino acid sequence homology with the respective nucleotide or amino acid sequence of the human coronavirus 229E or Porcine Epidemic Diarrhea Virus as for example can be found in Genbank (for example in accession number AF304460 (HCoV 229E) or AF353511 (PEDV).

Sequence divergence of EMCR-CoV strains around the world may be somewhat higher, in analogy with other coronaviruses.

The term "nucleotide sequence homology" as used herein denotes the presence of homology between two (poly)nucleotides. Polynucleotides have "homologous" sequences if the sequence of nucleotides in the two sequences is the same when aligned for maximum correspondence. Sequence comparison between two or more polynucleotides is generally performed by comparing portions of the two sequences over a comparison window to identify and compare local regions of sequence similarity. The comparison window is generally from about 20 to 200 contiguous nucleotides. The "percentage of sequence homology" for polynucleotides, such as 50, 60, 70, 80, 90, 95, 98, 99 or 100 percent sequence homology may be determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide

sequence in the comparison window may include additions or deletions (i.e. gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by: (a) determining the number of positions at which the identical nucleic acid base occurs in both sequences to yield the number of matched positions; (b) dividing the number of matched positions by the total number of positions in the window of comparison; and (c) multiplying the result by 100 to yield the percentage of sequence homology. Optimal alignment of sequences for comparison may be conducted by computerized implementations of known algorithms, or by inspection. Readily available sequence comparison and multiple sequence alignment algorithms are, respectively, the Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. 1990. J. Mol. Biol. 215:403; Altschul, S.F. et al. 1997. Nucleic Acid Res. 25:3389-3402) and ClustalW programs both available on the internet. Other suitable programs include GAP, BESTFIT and FASTA in the Wisconsin Genetics Software Package (Genetics Computer Group (GCG), Madison, WI, USA).

As used herein, "substantially complementary" means that two nucleic acid sequences have at least about 65%, preferably about 70%, more preferably about 80%, even more preferably 90%, and most preferably about 98%, sequence complementarity to each other. This means that the primers and probes must exhibit sufficient complementarity to their template and target nucleic acid, respectively, to hybridise under stringent conditions. Therefore, the primer sequences as disclosed in this specification need not reflect the exact sequence of the binding region on the template and degenerate primers can be used. A substantially complementary primer sequence is one that has sufficient sequence complementarity to the amplification template to result in primer binding and second-strand synthesis.

The term "hybrid" refers to a double-stranded nucleic acid molecule, or duplex, formed by hydrogen bonding between complementary nucleotides. The terms "hybridise" or "anneal" refer to the process by which single strands of nucleic acid sequences form double-helical segments through hydrogen bonding between complementary nucleotides

The term "oligonucleotide" refers to a short sequence of nucleotide monomers (usually 6 to 100 nucleotides) joined by phosphorous linkages (e.g., phosphodiester, alkyl and aryl-phosphate, phosphorothioate), or non-phosphorous linkages (e.g., peptide, sulfamate and others). An oligonucleotide may contain modified nucleotides having modified bases (e.g., 5-methyl cytosine) and modified sugar groups (e.g., 2'-O-methyl

ribosyl, 2'-O-methoxyethyl ribosyl, 2'-fluoro ribosyl, 2'-amino ribosyl, and the like). Oligonucleotides may be naturally-occurring or synthetic molecules of double- and single-stranded DNA and double- and single-stranded RNA with circular, branched or linear shapes and optionally including domains capable of forming stable secondary structures (e.g., stem-and-loop and loop-stem-loop structures).

The term "primer" as used herein refers to an oligonucleotide which is capable of annealing to the amplification target allowing a DNA polymerase to attach thereby serving as a point of initiation of DNA synthesis when placed under conditions in which synthesis of primer extension product which is complementary to a nucleic acid strand is induced, i.e., in the presence of nucleotides and an agent for polymerization such as DNA polymerase and at a suitable temperature and pH. The (amplification) primer is preferably single stranded for maximum efficiency in amplification. Preferably, the primer is an oligodeoxy ribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the agent for polymerization. The exact lengths of the primers will depend on many factors, including temperature and source of primer. A "pair of bi-directional primers" as used herein refers to one forward and one reverse primer as commonly used in the art of DNA amplification such as in PCR amplification.

The term "probe" refers to a single-stranded oligonucleotide sequence that will recognize and form a hydrogen-bonded duplex with a complementary sequence in a target nucleic acid sequence analyte or its cDNA derivative.

The terms "stringency" or "stringent hybridization conditions" refer to hybridization conditions that affect the stability of hybrids, e.g., temperature, salt concentration, pH, formamide concentration and the like. These conditions are empirically optimised to maximize specific binding and minimize non-specific binding of primer or probe to its target nucleic acid sequence. The terms as used include reference to conditions under which a probe or primer will hybridise to its target sequence, to a detectably greater degree than other sequences (e.g. at least 2-fold over background). Stringent conditions are sequence dependent and will be different in different circumstances. Longer sequences hybridise specifically at higher temperatures. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridises to a perfectly matched probe or primer.

Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M Na<sup>+</sup> ion, typically about 0.01 to 1.0 M Na<sup>+</sup> ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes or primers (e.g. 10 to 50 nucleotides) and at least about 60°C for long probes or primers (e.g. greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary low stringent conditions or "conditions of reduced stringency" include hybridization with a buffer solution of 30% formamide, 1 M NaCl, 1% SDS at 37°C and a wash in 2x SSC at 40°C. Exemplary high stringency conditions include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1x SSC at 60°C. Hybridization procedures are well known in the art and are described in e.g. Ausubel et al, *Current Protocols in Molecular Biology*, John Wiley & Sons Inc., 1994.

The term "antibody" includes reference to antigen binding forms of antibodies (e. g., Fab, F (ab) 2). The term "antibody" frequently refers to a polypeptide substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof which specifically bind and recognize an analyte (antigen). However, while various antibody fragments can be defined in terms of the digestion of an intact antibody, one of skill will appreciate that such fragments may be synthesized de novo either chemically or by utilizing recombinant DNA methodology. Thus, the term antibody, as used herein, also includes antibody fragments such as single chain Fv, chimeric antibodies (i. e., comprising constant and variable regions from different species), humanized antibodies (i. e., comprising a complementarity determining region (CDR) from a non-human source) and heteroconjugate antibodies (e. g., bispecific antibodies).

In short, the invention provides an isolated essentially mammalian positive-sense single stranded RNA virus (EMCR-CoV) belonging to the Coronaviruses and identifiable as phylogenetically corresponding thereto by determining a nucleic acid sequence of a suitable fragment of the genome of said virus and testing it in phylogenetic tree analyses wherein maximum likelihood trees are generated using 100 bootstraps and 3 jumbles and finding it to be more closely phylogenetically corresponding to a virus isolate having the sequences as depicted in figure 1 than it is corresponding to a virus isolate of PEDV (porcine epidemic diarrhea virus), HCoV-229E (human coronavirus 229E), PRCoV (porcine respiratory coronavirus), TGEV (transmissible gastroenteritis virus), CaCoV (Canine coronavirus) and FeCoV (feline coronavirus).

Suitable nucleic acid genome fragments each useful for such phylogenetic tree analyses are for example any of the fragments encoding the Matrix protein or the Nucleocapsid protein as disclosed in Figure 1, leading to the phylogenetic tree analysis as disclosed herein in figure 2a or 2b. Other suitable nucleic acid fragments useful for such phylogenetic tree analyses are for example any of the fragments encoding Replicase 1a and 1b, Spike, orf 4a and 4b, and E.

A suitable open reading frame (ORF) useful in phylogenetic analyses comprises the ORF encoding the viral replicase (ORF 1a). When an overall amino acid identity of at least 60%, preferably of at least 70%, more preferably of at least 80%, more preferably of at least 90%, most preferably of at least 95% of the analysed replicase with the replicase having a sequence comprising the amino acids of Figure 1 is found, the analysed virus isolate comprises an EMCR-CoV virus isolate according to the invention.

A suitable open reading frame (ORF) useful in phylogenetic analyses comprises the ORF encoding the viral replicase (ORF 1b). When an overall amino acid identity of at least 82%, more preferably of at least 90%, most preferably of at least 95% of the analysed replicase with the replicase having a sequence comprising the amino acids of Figure 1 is found, the analysed virus isolate comprises an EMCR-CoV virus isolate according to the invention.

Another suitable open reading frame (ORF) useful in phylogenetic analyses comprises the ORF encoding the Nucleocapsid protein. When an overall amino acid identity of at least 50%, more preferably of at least 60%, more preferably of at least 70%, more preferably of at least 80%, more preferably of at least 90%, most preferably of at least 95% of the analysed Nucleocapsid protein with the Nucleocapsid protein encoded by a sequence comprising (part of) the sequence F of Figure 1 is found, the analysed virus isolate comprises an EMCR-CoV isolate according to the invention.

Another suitable open reading frame (ORF) useful in phylogenetic analyses comprises the ORF encoding the Matrix protein. When an overall amino acid identity of at least 60%, more preferably of at least 70%, more preferably of at least 80%, more preferably of at least 90%, most preferably of at least 95% of the analysed Matrix protein with the Matrix protein encoded by a sequence comprising (part of) the sequence F of Figure 1 is found, the analysed virus isolate comprises an EMCR-CoV isolate according to the invention.

Another suitable open reading frame (ORF) useful in phylogenetic analyses comprises the ORF encoding the spike protein S. When an overall amino acid identity of

at least 55%, more preferably of at least 60%, more preferably of at least 70%, more preferably of at least 80%, more preferably of at least 90%, most preferably of at least 95% of the analysed S-protein encoded by a sequence comprising the sequence of translation 2 of E and translation 1 of the F sequence of the S-protein as depicted in Figure 1 is found, the analysed virus isolate comprises an EMCR-CoV virus isolate according to the invention. The S ORF of the EMCR-CoV virus seems to be located adjacent to the ORF 1ab (coding for the viral replicase), which would discriminate an EMCR-CoV viruses from the bovine coronavirus and the murine hepatitis virus, which have a so-called 2a gene and an HE-gene between the S protein and the viral polymerase.

The invention provides among others an isolated or recombinant nucleic acid or virus-specific functional fragment thereof obtainable from a virus according to the invention. The isolated or recombinant nucleic acids comprises the sequences as given in figure 1 or sequences of homologues which are able to hybridise with those under stringent conditions. In particular, the invention provides primers and/or probes suitable for identifying an EMCR-CoV virus nucleic acid.

Furthermore, the invention provides a vector comprising a nucleic acid according to the invention. To begin with, vectors such as plasmid vectors containing (parts of) the genome of the EMCR-CoV virus, virus vectors containing (parts of) the genome of the EMCR-CoV (for example, but not limited thereto, vaccinia virus, retroviruses, baculovirus), or EMCR-CoV virus containing (parts of) the genome of other virus or other pathogens are provided.

Also, the invention provides a host cell comprising a nucleic acid or a vector according to the invention. Plasmid or viral vectors containing the replicase components of EMCR-CoV virus are generated in prokaryotic cells for the expression of the components in relevant cell types (bacteria, insect cells, eukaryotic cells). Plasmid or viral vectors containing full-length or partial copies of the EMCR-CoV virus genome will be generated in prokaryotic cells for the expression of viral nucleic acids *in-vitro* or *in-vivo*. The latter vectors may contain other viral sequences for the generation of chimeric viruses or chimeric virus proteins, may lack parts of the viral genome for the generation of replication defective virus, and may contain mutations, deletions or insertions for the generation of attenuated viruses.



Infectious copies of EMCR-CoV virus (being wild type, attenuated, replication-defective or chimeric) can be produced upon co-expression of the polymerase component according to the state-of-the-art technologies described above.

In addition, eukaryotic cells, transiently or stably expressing one or more full-length or partial EMCR-CoV virus proteins can be used. Such cells can be made by transfection (proteins or nucleic acid vectors), infection (viral vectors) or transduction (viral vectors) and may be useful for complementation of mentioned wild type, attenuated, replication-defective or chimeric viruses.

A chimeric virus may be of particular use for the generation of recombinant vaccines protecting against two or more viruses. For example, it can be envisaged that EMCR-CoV virus vector expressing one or more proteins of a human metapneumovirus or a human metapneumovirus vector expressing one or more proteins of EMCR-CoV virus will protect individuals vaccinated with such vector against both virus infections. Such a specific chimeric virus is particularly useful in the invention because it is suspected that co-infection of, for instance, human metapneumovirus frequently occurs in coronavirus infected patients. Attenuated and replication-defective viruses may be used for vaccination purposes with live vaccines as has been suggested for other viruses.

In a preferred embodiment, the invention provides a proteinaceous molecule or coronavirus-specific viral protein or functional fragment thereof encoded by a nucleic acid according to the invention. Useful proteinaceous molecules are for example derived from any of the genes or genomic fragments derivable from a virus according to the invention. Such molecules, or antigenic fragments thereof, as provided herein, are for example useful in diagnostic methods or kits and in pharmaceutical compositions such as sub-unit vaccines and inhibitory peptides. Particularly useful are the viral replicase protein, the spike protein, the matrix protein, the nucleocapsid or antigenic fragments thereof for inclusion as antigen or subunit immunogen, but inactivated whole virus can also be used. Particularly useful are also those proteinaceous substances that are encoded by recombinant nucleic acid fragments that are identified for phylogenetic analyses, of course preferred are those that are within the preferred bounds and metes of ORFs useful in phylogenetic analyses, in particular for eliciting EMCR-CoV virus specific antibodies, whether in vivo (e.g. for protective purposes or for providing diagnostic antibodies) or in vitro (e.g. by phage display technology or another technique useful for generating synthetic antibodies).

Also provided herein are antibodies, be it natural polyclonal or monoclonal, or synthetic (e.g. (phage) library-derived binding molecules) antibodies that specifically react with an antigen comprising a proteinaceous molecule or EMCR-CoV virus-specific functional fragment thereof according to the invention. Such antibodies are useful in a method for identifying a viral isolate as an EMCR-CoV virus comprising reacting said viral isolate or a component thereof with an antibody as provided herein. This can for example be achieved by using purified or non-purified EMCR-CoV virus or parts thereof (proteins, peptides) using ELISA, RIA, FACS or similar formats of antigen detection assays (Current Protocols in Immunology). Alternatively, infected cells or cell cultures may be used to identify viral antigens using classical immunofluorescence or immunohistochemical techniques. Specifically useful in this respect are antibodies raised against EMCR-CoV virus proteins which are encoded by a nucleotide sequence comprising one or more of the sequences disclosed in figure 1.

Other methods for identifying a viral isolate as an EMCR-CoV virus comprise reacting said viral isolate or a component thereof with a virus specific nucleic acid according to the invention.

In this way the invention provides a viral isolate identifiable with a method according to the invention as a mammalian virus taxonomically corresponding to a positive-sense single stranded RNA virus identifiable as likely belonging to the EMCR-CoV virus genus within the family of Coronaviruses.

The method is useful in a method for virologically diagnosing an EMCR-CoV virus infection of a mammal, said method for example comprising determining in a sample of said mammal the presence of a viral isolate or component thereof by reacting said sample with a nucleic acid or an antibody according to the invention.

Methods of the invention can in principle be performed by using any nucleic acid amplification method, such as the Polymerase Chain Reaction (PCR; Mullis 1987, U.S. Pat. No. 4,683,195, 4,683,202, en 4,800,159) or by using amplification reactions such as Ligase Chain Reaction (LCR; Barany 1991, Proc. Natl. Acad. Sci. USA 88:189-193; EP Appl. No., 320,308), Self-Sustained Sequence Replication (3SR; Guatelli et al., 1990, Proc. Natl. Acad. Sci. USA 87:1874-1878), Strand Displacement Amplification (SDA; U.S. Pat. Nos. 5,270,184, en 5,455,166), Transcriptional Amplification System (TAS; Kwoh et al., Proc. Natl. Acad. Sci. USA 86:1173-1177), Q-Beta Replicase (Lizardi et al., 1988, Bio/Technology 6:1197), Rolling Circle Amplification (RCA; U.S. Pat. No. 5,871,921), Nucleic Acid Sequence Based Amplification (NASBA), Cleavase Fragment

Length Polymorphism (U.S. Pat. No. 5,719,028), Isothermal and Chimeric Primer-initiated Amplification of Nucleic Acid (ICAN), Ramification-extension Amplification Method (RAM; U.S. Pat. Nos. 5,719,028 and 5,942,391) or other suitable methods for amplification of nucleic acids.

- 5 In order to amplify a nucleic acid with a small number of mismatches to one or more the amplification primers, an amplification reaction may be performed under conditions of reduced stringency (e.g. a PCR amplification using an annealing temperature of 38°C, or the presence of 3.5 mM MgCl<sub>2</sub>). The person skilled in the art will be able to select conditions of suitable stringency.

- 10 The primers herein are selected to be "substantially" complementary (i.e. at least 65%, more preferably at least 80% perfectly complementary) to their target regions present on the different strands of each specific sequence to be amplified. It is possible to use primer sequences containing e.g. inositol residues or ambiguous bases or even primers that contain one or more mismatches when compared to the target sequence  
15 general, sequences that exhibit at least 65%, more preferably at least 80% homology with the target DNA or RNA oligonucleotide sequences, are considered suitable for use in a method of the present invention. Sequence mismatches are also not critical when using low stringency hybridization conditions.

- The detection of the amplification products can in principle be accomplished by  
20 any suitable method known in the art. The detection fragments may be directly stained or labelled with radioactive labels, antibodies, luminescent dyes, fluorescent dyes, or enzyme reagents. Direct DNA stains include for example intercalating dyes such as acridine orange, ethidium bromide, ethidium monoazide or Hoechst dyes.

- Alternatively, the DNA or RNA fragments may be detected by incorporation of labelled  
25 dNTP bases into the synthesized fragments. Detection labels which may be associated with nucleotide bases include e.g. fluorescein, cyanine dye or BrdUrd.

- When using a probe-based detection system, a suitable detection procedure for use in the present invention may for example comprise an enzyme immunoassay (EIA) format (Jacobs et al., 1997; J. Clin. Microbiol. 35, 791-795). For performing a detection  
30 by manner of the EIA procedure, either the forward or the reverse primer used in the amplification reaction may comprise a capturing group, such as a biotin group for immobilization of target DNA PCR amplicons on e.g. a streptavidin coated microtiter plate wells for subsequent EIA detection of target DNA -amplicons (see below). The

skilled person will understand that other groups for immobilization of target DNA PCR amplicons in an EIA format may be employed.

Probes useful for the detection of the target DNA as disclosed herein preferably bind only to at least a part of the DNA sequence region as amplified by the DNA amplification procedure. Those of skill in the art can prepare suitable probes for detection based on the nucleotide sequence of the target DNA without undue experimentation as set out herein. Also the complementary nucleotide sequences, whether DNA or RNA or chemically synthesized analogs, of the target DNA may suitably be used as type-specific detection probes in a method of the invention, provided that such a complementary strand is amplified in the amplification reaction employed.

Suitable detection procedures for use herein may for example comprise immobilization of the amplicons and probing the DNA sequences thereof by e.g. southern blotting. Other formats may comprise an EIA format as described above. To facilitate the detection of binding, the specific amplicon detection probes may comprise a label moiety such as a fluorophore, a chromophore, an enzyme or a radio-label, so as to facilitate monitoring of binding of the probes to the reaction product of the amplification reaction. Such labels are well-known to those skilled in the art and include, for example fluorescein isothiocyanate (FITC),  $\beta$ -galactosidase, horseradish peroxidase, streptavidin, biotin, digoxigenin,  $^{35}\text{S}$  or  $^{125}\text{I}$ . Other examples will be apparent to those skilled in the art.

Detection may also be performed by a so called reverse line blot (RLB) assay, such as for instance described by Van den Brule et al. (2002, J. Clin. Microbiol. 40, 779-787). For this purpose RLB probes are preferably synthesized with a 5' amino-group for subsequent immobilization on e.g. carboxyl-coated nylon membranes. The advantage of an RLB format is the ease of the system and its speed, thus allowing for high throughput sample processing.

The use of nucleic acid probes for the detection of RNA or DNA fragments is well known in the art. Mostly these procedure comprise the hybridization of the target nucleic acid with the probe followed by post-hybridization washings. Specificity is typically the function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For nucleic acid hybrids, the  $T_m$  can be approximated from the equation of Meinkoth and Wahl, Anal. Biochem., 138: 267-284 (1984):  $T_m = 81.5\text{ }^\circ\text{C} + 16.6 (\log M) + 0.41 (\% \text{ GC}) - 0.61 (\% \text{ form}) - 500/L$ ; where  $M$  is the molarity of monovalent cations, % GC is the percentage of guanosine and cytosine

nucleotides in the nucleic acid, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridizes to a perfectly matched probe.  $T_m$  is reduced by about 1 °C each 1 % of mismatching; thus, the hybridization and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with > 90% identity are sought, the  $T_m$  can be decreased 10°C. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point ( $T_m$ ) for the specific sequence and its complement at a defined ionic strength and pH. However severely stringent conditions can utilize a hybridization and/or wash at 1,2,3, or 4 °C lower than the thermal melting point ( $T_m$ ); moderately stringent conditions can utilize hybridization and/or wash at 6, 7, 8, 9, or 10 °C lower than the thermal melting point ( $T_m$ ); low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 15, or 20 °C lower than the thermal melting point ( $T_m$ ). Using the equation, hybridization and wash compositions, and desired  $T_m$ , those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a  $T_m$  of less than 45 °C (aqueous solution) or 32 °C (formamide solution) it is preferred to increase the SSC concentration so that a higher temperature can be used. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Laboratory Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Acid Probes, Part I, Chapter 2* "Overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier. New York (1993); and *Current Protocols in Molecular Biology, Chapter 2*, Ausubel, et al., Eds., Greene Publishing and Wiley-Interscience, New York (1995).

In another aspect, the invention provides oligonucleotide probes for the generic detection of target RNA or DNA. The detection probes herein are selected to be "substantially" complementary to one of the strands of the double stranded nucleic acid generated by an amplification reaction of the invention. Preferably the probes are substantially complementary to the immobilizable, e.g. biotin labelled, antisense strand of the amplicons generated from the target RNA or DNA.

It is allowable for detection probes of the present invention to contain one or more mismatches to their target sequence. In general, sequences that exhibit at least 65%, more preferably at least 80% homology with the target oligonucleotide sequences are considered suitable for use in a method of the present invention.

Antibodies, both monoclonal and polyclonal, can also be used for detection purpose in the present invention, for example, in immunoassays in which they can be utilized in liquid phase or bound to a solid phase carrier. In addition, the monoclonal antibodies in these immunoassays can be detectably labeled in various ways. A variety  
5 of immunoassay formats may be used to select antibodies specifically reactive with a particular protein (or other analyte). For example, solid-phase ELISA immunoassays are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See Harlow and Lane, *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, New York (1988), for a description of immunoassay formats and conditions  
10 that can be used to determine selective binding. Examples of types of immunoassays that can utilize antibodies of the invention are competitive and non-competitive immunoassays in either a direct or indirect format. Examples of such immunoassays are the radioimmunoassay (RIA) and the sandwich (immunometric) assay. Detection of the antigens using the antibodies of the invention can be done utilizing immunoassays that  
15 are run in either the forward, reverse, or simultaneous modes, including immunohistochemical assays on physiological samples. Those of skill in the art will know, or can readily discern, other immunoassay formats without undue experimentation.

Antibodies can be bound to many different carriers and used to detect the  
20 presence of the target molecules. Examples of well-known carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, agaroses and magnetite. The nature of the carrier can be either soluble or insoluble for purposes of the invention. Those skilled in the art will know of other suitable carriers for binding monoclonal antibodies, or will be able to  
25 ascertain such using routine experimentation.

The invention also provides a method for serologically diagnosing an EMCR-CoV virus infection of a mammal comprising determining in a sample of said mammal the presence of an antibody specifically directed against an EMCR-CoV virus or component thereof by reacting said sample with a proteinaceous molecule or fragment thereof or an  
30 antigen according to the invention

Methods and means provided herein are particularly useful in a diagnostic kit for diagnosing an EMCR-CoV virus infection, be it by virological or serological diagnosis. Such kits or assays may for example comprise a virus, a nucleic acid, a proteinaceous molecule or fragment thereof, an antigen and/or an antibody according to the invention

Use of a virus, a nucleic acid, a proteinaceous molecule or fragment thereof, an antigen and/or an antibody according to the invention is also provided for the product of a pharmaceutical composition, for example for the treatment or prevention of EMC-CoV virus infections and/or for the treatment or prevention of atypical pneumonia, in particular in humans. Preferably a peptide comprising part of the amino acid sequence of the spike protein as depicted in the relevant translations of Figure 1, is used for the preparation of a therapeutic or prophylactic peptide. Also preferably, a protein comprising the amino acid sequence of the spike protein as depicted in the relevant translations of Figure 1, is used for the preparation of a sub-unit vaccine. Furthermore the nucleocapsid of Coronaviruses, as depicted in the translation of Figure 1, is known to be particularly useful for eliciting cell-mediated immunity against Coronaviruses and can be used for the preparation of a sub-unit vaccine.

Attenuation of the virus can be achieved by established methods developed for this purpose, including but not limited to the use of related viruses of other species, serial passages through laboratory animals or/and tissue/cell cultures, serial passages through cell cultures at temperatures below 37°C (cold-adaptation), site directed mutagenesis of molecular clones and exchange of genes or gene fragments between related viruses.

A pharmaceutical composition comprising a virus, a nucleic acid, a proteinaceous molecule or fragment thereof, an antigen and/or an antibody according to the invention can for example be used in a method for the treatment or prevention of an EMC-CoV virus infection and/or a respiratory illness comprising providing an individual with a pharmaceutical composition according to the invention. This is most useful when said individual comprises a human. Antibodies against EMC-CoV virus proteins, especially against the spike protein of EMC-CoV virus, preferably against the amino acid sequence as depicted in the translation in figure 1, are also useful for prophylactic or therapeutic purposes, as passive vaccines. It is known from other coronaviruses that the spike protein is a very strong antigen and that antibodies against spike protein can be used in prophylactic and therapeutic vaccination.

The invention also provides method to obtain an antiviral agent useful in the treatment of atypical pneumonia comprising establishing a cell culture or experimental animal comprising a virus according to the invention, treating said culture or animal with an candidate antiviral agent, and determining the effect of said agent on said virus or its infection of said culture or animal. An example of such an antiviral agent

comprises an EMCR-CoV virus-neutralising antibody, or functional component thereof, as provided herein, but antiviral agents of other nature are obtained as well.

5 The invention also provides use of an antiviral agent according to the invention for the preparation of a pharmaceutical composition, in particular for the preparation of a pharmaceutical composition for the treatment of atypical pneumonia, specifically when caused by an EMCR-CoV virus infection, and provides a pharmaceutical composition comprising an antiviral agent according to the invention, useful in a method for the treatment or prevention of an EMCR-CoV virus infection or atypical pneumonia, said method comprising providing an individual with such a pharmaceutical  
10 composition.

The invention also comprises an animal model usable for testing of prophylactic and/or therapeutic methods and/or preparations. It is hypothesized that apes can be infected with the EMCR-CoV virus, thereby showing clinical symptoms, and more importantly, similar tissue morphology as found in humans suffering from atypical  
15 pneumonia caused by the EMCR-CoV virus. Subjecting apes to a prophylactic or therapeutic treatment either before or during infection with the virus will have a good and useful predictionary value for application of such a prophylaxis or therapy in human subjects.

The invention is further explained in the Examples without limiting it thereto.



## Figure legends

Fig. 1: Nucleotide sequences from parts of the EMCR-CoV virus. Also included are the putative amino acid sequences of polypeptides.

5

Fig. 2: Phylogenetic relationship for the nucleotide sequences of isolate EMCR-CoV with its closest relatives genetically. Phylogenetic trees were generated by maximum likelihood analyses using 100 bootstraps and 3 jumbles. The scale representing the number of nucleotide changes is shown for each tree. Figure 1a. Maximum likelihood tree of matrix gene nucleotide sequences. Numbers in trees represent bootstrap values. The scale bar roughly reflects 10 % nucleotide differences between related sequences. Figure 1b. Maximum likelihood tree of nucleocapsid gene nucleotide sequences. Numbers in trees represent bootstrap values. The scale bar roughly reflects 10 % nucleotide differences between related sequences.

10

Fig. 3: Similarity matrices indicating amino acid identity for the putative Replicase 1a, Replicase 1b, Replicase 1ab, Spike, Orf E, Matrix and Nucleocapsid proteins (3a-g, respectively), and for the putative Matrix protein and Nucleoprotein (3h and 3i resp.) between the EMCR-CoV virus and closely related coronaviruses. See text for abbreviations.

15

Figure 4 Alignments with various coronaviruses: 5'untranslated region genomic sequence (a); Putative orf 1a amino acid sequence (b); Putative orf 1b amino acid sequence (c); Putative orf 1ab amino acid sequence (d); Putative Spike amino acid sequence (e); Putative orf 4a amino acid sequence (f); Putative orf 4ab amino acid sequence (g); Putative orf E amino acid sequence (h); Putative Matrix amino acid sequence (i); Putative Nucleoprotein amino acid sequence (j); Putative 3'untranslated genomic sequence (k); See text for abbreviations.

20

25

## Examples

### *Specimen collection*

Virus was collected from an 8 month old patient suffering from pneumonia using nasal  
5 swabs.

### *Virus isolation and culture*

Throat swabs were dipped into a culture of tMK cells and passaged four times. Virus  
was then in Vero-118 cells. One litre of virus containing cell culture supernatant was  
10 harvested, and the virus was pelleted in an ultracentrifuge and the virus pellet was  
resuspended in 1ml PBS.

### *RNA isolation*

RNA was isolated from the supernatant of infected cell cultures or sucrose gradient  
15 fractions using a High Pure RNA Isolation kit according to instructions from the  
manufacturer (Roche Diagnostics, Almere, The Netherlands).

### *Sequencing*

Purified RNA was sent to BaseClear holding BV (Leiden, The Netherlands) for  
20 sequencing.

### *Phylogenetic analyses*

Nucleotide sequences were aligned using Clustal W running under BioEdit version  
5.0.9. Maximum likelihood trees were created using the Seqboot and DNA-ML packages  
25 of Phylip 5.6 using 100 bootstraps and 3 jumbles. The consensus trees were calculated  
using the Consense package of phylip 5.6. These consensus trees were used as usertree  
in DNA-ML to recalculate the branch lengths from the original sequences.

The sequences of EMCR-CoV were compared with those of reference viruses  
30 representing each species in the four groups of coronaviruses. These were: human  
coronavirus 229E (229E), af304460; porcine epidemic diarrhea virus (PEDV) af353511;  
transmissible gastroenteritis virus (TGEV), aj271965; bovine coronavirus (BoCoV),  
af220295; murine hepatitis virus (MHV), af201929; avian infectious bronchitis virus  
(AIBV), m95169, Canine coronavirus (CaCoV), d13096; feline coronavirus (FeCoV),

ay204704; porcine respiratory coronavirus (PRCoV), z24675; human coronavirus OC4 (OC43), m76373, l14643, m933990; porcine haemagglutinating encephalomyelitis virus (HEV), ay078417; rat coronavirus (RtCoV) af 207551) References for the viruses are the numbers of the NCBI catalog (<http://www.ncbi.nlm.nih.gov/entrez/>).

5

In general, coronaviruses, such as EMCRC-CoV can be isolated and identified according to the following protocol:

#### *Specimen collection*

In order to find virus isolates nasopharyngeal aspirates, throat and nasal swabs, broncho alveolar lavages, serum and plasma samples, and stools preferably from mammals such as humans, carnivores (dogs, cats, mustelids, seals etc.), horses, ruminants (cattle, sheep, goats etc.), pigs, rabbits, birds (poultry, ostriches, etc) should be examined. From birds cloaca swabs and droppings can be examined as well. Sera should be collected for immunological assays, such as ELISA, molecular-based assays, such as RT-PCR and virus neutralisation assays.

15

Collected virus specimens may be diluted with 5 ml Dulbecco MEM medium (BioWhittaker, Walkersville, MD) and thoroughly mixed on a vortex mixer for one minute. The suspension is thus centrifuged for ten minutes at 840 x g. The sediment is spread on a multispot slide (Nutacon, Leimuiden, The Netherlands) for immunofluorescence techniques, and the supernatant is used for virus isolation.

20

#### *Virus isolation*

For virus isolation Vero-118 cells or tMK cells (RIVM, Bilthoven, The Netherlands) were cultured in 24 well plates containing glass slides (Costar, Cambridge, UK), with the medium described below supplemented with 10% fetal bovine serum (BioWhittaker, Vervier, Belgium). Before inoculation the plates were washed with PBS and supplied with Eagle's MEM with Hanks' salt (ICN, Costa mesa, CA) supplemented with 0.52/lit gram  $\text{NaHCO}_3$ , 0.025 M Hepes (Biowhittaker), 2 mM L-glutamine (Biowhittaker), 200 units/liter penicilline, 200  $\mu\text{g/liter}$  streptomycine (Biowhittaker), 1gram/liter lactalbumine (Sigma-Aldrich, Zwijndrecht, The Netherlands), 2.0 gram/liter D-glucose (Merck, Amsterdam, The Netherlands), 10 gram/liter peptone (Oxoid, Haarlem, The Netherlands) and 0.02% trypsin (Life Technologies, Bethesda, MD). The plates were inoculated with supernatant of the patient samples, 0,2 ml per well in triplicate, followed by centrifuging at 840x g for one hour. After inoculation the plates were

30

incubated at 37 °C for 1-7 days and cultures were checked daily for CPE. Extensive CPE was generally observed within 5-10 and included detachment of cells from the monolayer..

5 *Virus culture*

Sub-confluent monolayers of tMK cells or Vero clone 118 cells in media as described above were inoculated with supernatants of samples that displayed CPE or with samples taken from a patient.

10 *RNA isolation*

RNA was isolated from the supernatant of infected cell cultures or sucrose gradient fractions using a High Pure RNA Isolation kit according to instructions from the manufacturer (Roche Diagnostics, Almere, The Netherlands). RNA can also be isolated following other procedures known in the field (*Current Protocols in Molecular Biology*).

15

*Sequence analysis*

Sequence analyses were performed by BaseClear holding BV (Leiden, The Netherlands)

20

01 DEC 2003

## Claims

112

1. An isolated essentially mammalian positive-sense single stranded RNA virus (EMCR-CoV) comprising the sequence of figure 1 or homologues thereof.  
5
2. An isolated positive-sense single stranded RNA virus (EMCR-CoV) belonging to the Coronaviruses and identifiable as phylogenetically corresponding thereto by determining a nucleic acid sequence of said virus and testing it in phylogenetic tree analyses wherein maximum likelihood trees are generated using 100 bootstraps and 3  
10 jumbles and finding it to be more closely phylogenetically corresponding to a virus isolate having the sequences as depicted in figure 1 than it is corresponding to a virus isolate of PEDV (porcine epidemic diarrhea virus), HCoV-229E (human coronavirus 229E), PRCoV (porcine respiratory coronavirus), TGEV (transmissible gastroenteritis virus), CaCoV (Canine coronavirus) and FeCoV (feline coronavirus).  
15
3. A virus according to claim 1 or 2 wherein said nucleic acid sequence comprises an open reading frame (ORF) encoding a viral protein of said virus.
4. A virus according to claim 3 wherein said open reading frame is selected from the  
20 group of ORFs encoding the viral replicase, nuclear capsid protein, matrix protein and the spike protein.
5. A virus according to claim 1-4 isolatable from a human with atypical pneumonia.
- 25 6. An isolated or recombinant nucleic acid or EMCRCoV virus-specific functional fragment thereof obtainable from a virus according to anyone of claims 1 to 5.
7. A vector comprising a nucleic acid according to claim 6.
- 30 8. A host cell comprising a nucleic acid according to claim 6 or a vector according to claim 7.
9. An isolated or recombinant proteinaceous molecule or EMCRCoV virus-specific functional fragment thereof encoded by a nucleic acid according to claim 6.

10. An antigen comprising a proteinaceous molecule or EMCR-CoV virus-specific functional fragment thereof according to claim 9.
- 5 11. An antibody specifically directed against an antigen according to claim 10.
12. A method for identifying a viral isolate as an EMCR-CoV virus comprising reacting said viral isolate or a component thereof with an antibody according to claim 11.
- 10 13. A method for identifying a viral isolate as an EMCR-CoV virus comprising reacting said viral isolate or a component thereof with a nucleic acid according to claim 6.
- 15 14. A method for virologically diagnosing an EMCR-CoV infection of a mammal comprising determining in a sample of said mammal the presence of a viral isolate or component thereof by reacting said sample with a nucleic acid according to claim 6 or an antibody according to claim 11.
- 20 15. A method for serologically diagnosing an EMCR-CoV infection of a mammal comprising determining in a sample of said mammal the presence of an antibody specifically directed against an EMCR-CoV virus or component thereof by reacting said sample with a proteinaceous molecule or fragment thereof according to claim 9 or an antigen according to claim 10.
- 25 16. A diagnostic kit for diagnosing an EMCR-CoV infection comprising a virus according to anyone of claims 1 to 5, a nucleic acid according to claim 6, a proteinaceous molecule or fragment thereof according to claim 9, an antigen according to claim 10 and/or an antibody according to claim 11.
- 30 17. Use of a virus according to any one claims 1 to 5, a nucleic acid according to claim 6, a vector according to claim 7, a host cell according to claim 8, a proteinaceous molecule or fragment thereof according to claim 9, an antigen according to claim 10, or an antibody according to claim 11 for the production of a pharmaceutical composition.

18. Use according to claim 17 for the production of a pharmaceutical composition for the treatment or prevention of an EMCRC-CoV virus infection.
- 5 19. Use according to claim 17 or 18 for the production of a pharmaceutical composition for the treatment or prevention of atypical pneumonia.
20. A pharmaceutical composition comprising a virus according to any one of claims 1 to 5, a nucleic acid according to claim 6, a vector according to claim 7, a host cell  
10 according to claim 8, a proteinaceous molecule or fragment thereof according to claim 9, an antigen according to claim 10, or an antibody according to claim 11.
21. A method for the treatment or prevention of an EMCRC-CoV virus infection comprising providing an individual with a pharmaceutical composition according to  
15 claim 20.
22. A method for the treatment or prevention of atypical pneumonia comprising providing an individual with a pharmaceutical composition according to claim 20.
- 20 23. A viral replicase encoded by an RNA sequence comprising the indicated sequences, or homologues thereof as depicted in figure 1.
24. A viral spike protein comprising the indicated amino acid sequence as depicted in figure 1, or a homologue thereof.  
25
- 25 25. A viral nuclear capsid protein encoded by an RNA sequence comprising the indicated sequence as depicted in figure 1 or a homologue thereof.
26. A viral nsp 3 or envelope protein encoded by an RNA sequence comprising the  
30 indicated sequence as depicted in figure 1, or a homologue thereof.
27. A nucleic acid sequence which comprises one or more of the sequences coding for separate viral proteins as depicted in figure 1 or a nucleic acid sequence which can hybridise with any of these sequences under stringent conditions.

Abstract

112

5 The invention relates to the field of virology. The invention provides a new isolated essentially mammalian positive-sense single stranded RNA virus (EMCR-CoV) within the group of coronaviruses and components thereof.



01 December 2003 11:52

EMCR-CoV.MPD (1 > 27532) Site and Sequence

Enzymes : All 212 enzymes (No Filter)

Settings : Circular, Certain Sites Only, Standard Genetic Code

1/87

AGATAGAGAATTTTCTTATTTAGACTTTGTGTCTACTCCTCTCAACTAAACGAAATTTTCTAGTGCTGTCATTTGTTATGGCAGTCCTA  
TCTATCTCTTAAAAGAATAAATCTGAAACACAGATGAGGAGAGTTGATTTGCTTTAAAAAGATCACGACAGTAAACAATACCGTCAGGAT

5'UTR

GTAATTGAAATTTTCGTCAGTTTGTAACTGGTTAGGCAAGTGTGATTTTCTGTGTTTAAGCACTGGTGGTTCTGTCCACTAGTGCAC  
CATTAACTTTAAAGCAGTTCAAACATTTGACCAATCCGTTCAACATAAAAGACACAAATTCGTGACCACCAAGACAGGTGATCACGTG

5'UTR

ATTGATACTTAAGTGGTGTCTGTCACTGCTTATTGTGGAAGCAACGTTCTGTGCTTGTGGAACCAATAACTGCTAACCATGTTTTACA  
TAACTATGAATTCACCACAAGACAGTGACGAATAACACCTTCGTTGCAAGACAGCAACACCTTTGGTTATTGACGATTGGTACAAAATGT

5'UTR

M F Y  
L Replicase 1a

CAAGTGACACTTGCTGTTGCAAGTGATTCGGAATTTTCAGGTTTTGGTTTTGCCATTCTTCTGTAGCCGTTTCGCGCTTATAGCGAAGCCG  
GTTCACTGTGAACGACAACGTTCACTAAGCCTTTAAAGTCCAAAACCAAACGGTAAGGAAGACATCGGCAAGCGCAATATCGCTTCGGC

Q V T L A V A S D S E I S G F G F A I P S V A V R A Y S E A  
Replicase 1a

TGCACAAGGTTTTTCAGGCATGCCGCTTTGTTGCTTTTGGCTTACAGGATTGTGTAACCGGTATTAATGATGACGATTATGTCATTGCATTG  
ACGTGTTCCAAAAGTCCGTACGGCGAAACAACGAAACCGAATGTCCTAACACATTGGCCATAATTACTACTGCTAATACAGTAACGTAAC

A Q G F Q A C R F V A F G L Q D C V T G I N D D D Y V I A L  
Replicase 1a

CTGGTACTAATCAGCTTTGTGCCAAAATTTACTTTTTCTGATAGACCTCTTAATTTGCGAGGTTGGCTCATTTTTCTAACAGCAATTA  
GACCATGATTAGTCGAAACACGGTTTTAAATGAAAAAAGACTATCTGGAGAATTAACGCTCCAACCGAGTAAAAAAGATTGTCGTTAAT

T G T N Q L C A K I L L F S D R P L N L R G W L I F S N S N Y  
Replicase 1a

GTTCTTCAGGACTTTGATGTTGTTTTGGCCATGGTGCAGGAAGTGTGGTTTTGTGGATAAGTATATGTGTGGTTTTGATGGTAAACCTG  
CAAGAAGTCTGAAACTACAACAAAACCGGTACCACGTCCTTCACACAAAACACCTATTATATACACACAAAACCTACCATTGAC

V L Q D F D V V F G H G A G S V V F V D K Y M C G F D G K P  
Replicase 1a

GTTACCTAAAAACATGTGGGAATTTAGAGATTACTTTAATGATAATACTGATAGTATTGTTATTGGTGGTGTCACTTATCAATTAGCATGG  
CAATGGATTTTTGTACACCCTTAAATCTCTAATGAAATTACTATTGACTATCATAACAATAACCACCACAGTGAATAGTTAATCGTACCC

L P K N M W E F R D Y F N D N T D S I V I G G V T Y Q L A W  
Replicase 1a

ATGTTATACGTAAAGACCTTTCTTATGAACAGCAAAATGTTTTAGCTATTGAGAGCATTCAATTATCTTGGCACTACAGGTCATACTTTGAAC  
TACAATATGCATTTCTGGAAAGAATACTTGTGCTTTTACAAAATCGATAACTCTCGTAAGTAATAGAACCGTGATGTCCAGTATGAACTTC

D V I R K D L S Y E Q Q N V L A I E S I H Y L G T T G H T L K  
Replicase 1a

TC TGGTTG CAACTC ATTAATGCC AAGCCGCCTAA ATATTCTTCTA AGGTTG TTTTGAGTGGTGAATGGAATGCTGTGTATAAGGCGTTTGG  
AGACCAACGTTTGAGTAATTACGGTTGGGCGATTATAAGAAGATTCCAACAAACTCACCATTACCTTACGACACATATTCCGCAAACC

S G C K L I N A K P P K Y S S K V V L S G E W N A V Y K A F G  
Replicase 1a

TTCACCATTTATTACAAATGGTATATCATTGCTAGATATAATTGTTAAACCAGTTTTCTTTAATGCTTTTGTTAAATGCAATTGTGGTTCTG  
AAGTGGTAAATAATGTTTACCATATAGTAACGATCTATATTAACAATTTGGTCAAAAGAAATTACGAAAACAATTTACGTTAACACCAAGAC

S P F I T N G I S L L D I I V K P V F F N A F V K C N C G S  
Replicase 1a

AGAATTGGAGTGTGGTGCATGGGATGGTTATCTATCTTCTTGTGTGGCACACCTGCTAAGAACTTTGTGTTGTTCTCGGTAATGTTGTT  
TCTTAACCTCACAACCACGTACCCTACCAATAGATAGAAGAACAACACCGTGTGGACGATTCTTTGAAACACAACAAGGACCATTACAACAA

E N W S V G A W D G Y L S S C C G T P A K K L C V V P G N V V  
Replicase 1a

CCTGGTGATGTGATCATCACCTCAACTGATGCTGGTTGTGGTGTAAATACTATGCTGGCTTAGTTGTTAAACATATTACTAACATTACTGG  
GGACCACTACACTAGTAGTGAGTTGACTACGACCAACACCACAATTTATGATACGACCGAATCAACAATTTGTATAATGATTGTAATGACC

P G D V I I T S T D A G C G V K Y Y A G L V V K H I T N I T G  
Replicase 1a

TGTGTCTTTATGGCGTGTTACAGCTGTTTCTGATGGAATGTTTGTGGCAACATCTTCTTATGATGCACTTTTGCATAGAAATTCATTAG  
ACACAGAAATACCGCACAAATGTCGACAAGTAAGACTACCTTACAAACACCGTTGTAGAAGAATACTACGTGAAAACGTATCTTTAAGTAATC

V S L W R V T A V H S D G M F V A T S S Y D A L L H R N S L  
Replicase 1a

ACCCTTTTGCTTTGATGTTAACTTTACTTTCTAATCAATTACGTCTAGCTTTTCTTGGTGCTTCTGTTACAGAAGATGTTAAATTTGCT  
TGGGAAAACGAAACTACAATTGTGAAATGAAAGATTAGTTAATGCAGATCGAAAAGAACCACGAAGACAATGTCTTCTACAATTTAAACGA

D P F C F D V N T L L S N Q L R L A F L G A S V T E D V K F A  
Replicase 1a

GCTAGCACTGGTGTATTGACATTAGTGCTGGTATGTTTGGTCTTTACGATGACATATTGACAAACAATAAACCTTGGTTTGTACGCAAAGC  
CGATCGTGACCACAATAACTGTAATCACGACCATACAAACCAGAAATGCTACTGTATAACTGTTTGTATTGGAACCAACATGCGTTTCG

A S T G V I D I S A G M F G L Y D D I L T N N K P W F V R K A  
Replicase 1a

TTCTGGGCTTTTTGATGCAATCTGGGATGCTTTTGTGCGCTATTAAGCTTGTGCCAACTACTACTGGTGGTTTGGTTAGGTTTGTAAAGT  
AAGACCGGAAAAACTACGTTAGACCCTACGAAAACAACGGCGATAATTGAACACGGTTGATGATGACCACCAACCAATCCAACAATTCA

S G L F D A I W D A F V A A I K L V P T T T G G L V R F V K  
Replicase 1a

CTATCGCTTCAACTGTTTTAACTGTTTCTAATGGTGTATTATTATGTGTGCAGATGTTCCAGATGCTTTTCAACCAGTTTACCGCACATT  
GATAGCGAAGTTGACAAAATTGACAAAGATTACCACAATAATAACACACGTCTACAAGGTCTACGAAAAGTTGGTCAAATGGCGTGTAAG

S I A S T V L T V S N G V I I M C A D V P D A F Q P V Y R T F  
Replicase 1a

ACACAAGCTATTTGTGCTGCATTTGATTTTTCTTTAGATGTATTTAAAATTGGTGATGTTAAATTTAAACGACTTGGTGATTATGTTCTTA  
TGTGTTGATAAACACGACGTAACTAAAAAGAAATCTACATAAATTTTAACCACTACAATTTAAATTTGCTGAACCACTAATACAAGAAT

T Q A I C A A F D F S L D V F K I G D V K F K R L G D Y V L  
Replicase 1a

TGAAAATGCTCTTGTTCGTTTGACTACTGAAGTTGTTTCGTGGTGTTCGTGATGCTCGCATAAAGAAAGCCATGTTTACTAAAGTAGTTGTA  
ACTTTTACGAGAACAAGCAAAGTGAAGTCAACAAGCACCACAAGCACTACGAGCGTATTTCTTTTCGGTACAAATGATTTTCATCAACAT

E N A L V R L T T E V V R G V R D A R I K K A M F T K V V V  
Replicase 1a

GTCCTACAACTGAAGTTAAGTTTTCTGTTATTGAAGTTGCCACTGTTAATTTGCGTCTTGTGATTGTGCACCTGTAGTTTGCCCTAAAGG  
CAGGATGTTGACTTCAATTCAAAAGACAATAACTTGAACGGTGACAATTAAACGCAGAACAATAACACGTGGACATCAAACGGGATTTC

G P T T E V K F S V I E L A T V N L R L V D C A P V V C P K G  
Replicase 1a

AAAATTGTTGTTATTGCTGGACAAGCTTTTTCTATAGTGGTGGTTTTATCGTTTTATGGTTGATTCTACAAGTATTAAATGACCCTG  
TTTAAACAACAATAACGACCTGTTGAAAAAGATATCACCACCAAAAATAGCAAAATACCAACTAAGATGTTGACATAATTTACTGGGAC

K I V V I A G Q A F F Y S G G F Y R F M V D S T T V L N D P  
Replicase 1a

TTTTACTGGTGAGTTATTTTATACTATTAAGTTTAGTGGTTTTAAGCTTGATGGTTTTAACCATCAGTTTGTTAATGCTAGTTCTGCTACAC  
AAAATGACCACTCAATAAAATATGATAATTCAAATCACCAAAATTCGAACTACCAAAATTGGTAGTCAACAATTACGATCAAGACGATGTC

F T G E L F Y T I K F S G F K L D G F N H Q F V N A S S A T  
Replicase 1a

ATGCCATTATTGCTGTTGAGCTGTTGTTATCGGATTTTAAACTGCAGTTTTTGTGTACACATGTGTGGTTGATGGTTGTAGTGTCATTGTT  
TACGGTAATAACGACAACCTCGACAACAATAGCCTAAAATTTGACGTCAAAAACACATGTGTACACACCAACTACCAACATCACAGTAACAA

D A I I A V E L L L S D F K T A V F V Y T C V V D G C S V I V  
Replicase 1a

AGACGTGATGCTACATTCGCCACACATGTGTGTTTTAAGGACTGTTATAGTATTTGGGAGCAATTCTGCATTGATAATTGTGGTGAGCCATG  
TCTGCACTACGATGTAAGCGGTGTGTACACACAAAATTCCTGACAATATCATAAACCCCTCGTTAAGACGTAACATTAAACACCACTCGGTAC

R R D A T F A T H V C F K D C Y S I W E Q F C I D N C G E P V  
Replicase 1a

GTIIITGACTGATTATAATGCTATCTTGCAGAGTAATAACCCTCAATGTGCTATTGTTCAAGCATCGGAGTCTAAAGTTTTGCTTGAGAGGT  
CAAAAAGTACTAATATTACGATAGAACGTCTCATTATTGGGAGTTACACGATAACAAGTTCGTAGCCTCAGATTTCAAACGAACTCTCCA

F L T D Y N A I L Q S N N P Q C A I V Q A S E S K V L L E R  
Replicase 1a

TTTTACCTAAGTGCTCTGAAATACTGTTGAGTATTGATGATGGCCATTTATGGAATCTTTTTGTTGAAAAGTTTAATTTTGTACAGATTGG  
AAAATGGATTACAGGACTTTATGACAACCTATACTACTACCGGTAAATACCTTAGAAAAACAACCTTTCAAATTTAAACAATGTCTAACC

F L P K C P E I L L S I D D G H L W N L F V E K F N F V T D W  
Replicase 1a

TTAAAACTCTTAAGCTTACACTTACTTCTAATGGTCTTTTAGGTAATTGTGCCAAACGTTTTAGACGTGTTTTGGTAAAATTGCTTGATGT  
AATTTTGAAGAATTCGAATGTGAATGAAGATTACCAGAAAATCCATTAACACGGTTTGCAAATCTGCACAAAACCATTTTAACGAACTACA

L K T L K L T L T S N G L L G N C A K R F R R V L V K L L D V  
Replicase 1a

CTATAATGGTTTTCTTGAACTGTCTGTAGTGTCTGACACACTGCTGGTGTTCATTAAATATTATGCTGTTAATGTTCCATATGTAGTTA  
GATATTACCAAAAGAACTTTGACAGACATCACAGCATGTGTGACGACCACAAACGTAATTTATAATACGACAATTACAAGGTATACATCAAT

Y N G F L E T V C S V V H T A G V C I K Y Y A V N V P Y V V  
Replicase 1a

TTAGTGGTTTTGTAAGTCGTGTAATTCGTAGAGAAAGGTGTGACGTGACTTTTCCTTGTGTTAGTTGTGTCACCTTTTTCTATGAATTTTAA  
AATCACCAAAACATTACGACATTAAGCATCTCTTCCACACTGCACTGAAAAGGAACACAATCAACACAGTGAAAAAAGATACTTAAAAAT

I S G F V S R V I R R E R C D V T F P C V S C V T F F Y E F L  
Replicase 1a

GACACGTGTTTTGGTGTAGTAAACCTAATGCCATTGATGTTGAACATTTAGAGCTTAAAGAAACTGTTTTGTTGAACCTAAGGATGGTGG  
CTGTGCACAAAACCAATCATTGGATTACGGTAACTACAACCTGTAAATCTCGAATTTCTTGACAAAAACAACCTGGATTCTACCACC

D T C F G V S K P N A I D V E H L E L K E T V F V E P K D G G  
Replicase 1a

TCAATTTTTGTTTCTGATGATTATCTTTGGTATGTTGTAGATGACATTTATTATCCAGCTTCATGTAATGGTGTATTGCCAGTTGCTTTTA  
AGTTAAAAAACAAGACTACTAATAGAAACCATAACAACATCTACTGTAAATAATAGGTGGAAGTACATTACCACATAACGGTCAACGAAAAT

Q F F V S D D Y L W Y V V D D I Y Y P A S C N G V L P V A F  
Replicase 1a

CAAAATTGGCAGGTGGTAAATATCTTTTTCTGATGATGTTATAGTTCATGATGTTGAACCTACCCATAAAGTCAAGCTCATATTTGAGTTT  
GTTTAACCGTCCACCATTTTATAGAAAAAGACTACTACAATATCAAGTACTACAACCTGGATGGGTATTTTCAGTTTCAGTATAAACTCAAA

T K L A G G K I S F S D D V I V H D V E P T H K V K L I F E F  
Replicase 1a

GAAGATGATGTTGTTACCAGTCTTTGTAAGAAGAGTTTTGGTAAGTCTATTATTTATACAGGTGATTGGGAAGGTTTACATGAAGTTCTTAC  
CTTCTACTACAACAATGGTCAGAAACATTCTTCTCAAACCATTCAGATAATAAATATGTCCACTAACCCCTTCAAATGTACTTCAAGAATG

E D D V V T S L C K K S F G K S I I Y T G D W E G L H E V L T  
Replicase 1a

ATCTGCAATGAATGTCATTGGGCAACATATTAAGTTGCCACAATTTATATTTATGATGAAGAGGGTGGTTATGATGTTTCTAAACAGTTA  
TAGACGTTACTTACAGTAACCCGTTGTATAATTCAACGGTGTAAAAATATAAATACTACTTCTCCACCAATACTACAAAGATTTGGTCAAT

S A M N V I G Q H I K L P Q F Y I Y D E E G G Y D V S K P V  
Replicase 1a

TGATTTACAATGGCCTATTAGTGATGATGATGGTTGTGTTGTTGAAGCGAGCACTGATTTTCATCAATTAGAATCTGTTAGAGAAGAG  
ACTAAAGTGTTACGGATAATCACTACTATCACTACCAACACAACAACCTCGCTCGTGACTAAAAGTAGTTAATCTTAGACAATCTCTTCTC

M I S Q W P I S D D S D G C V V E A S T D F H Q L E S V R E E  
Replicase 1a

GTGATATAATTGAACAACCTTTTGGGGAAGTTGAACATGCGCTCTCAATTAGACAACCTTTTTCTTTTTCTTTAGAGATGAATTGGGTG  
CAACTATATTAACCTGTTGGAAAACCCCTTCAACTTGTACGCGAGAGTTAATCTGTTGGAAAAAGAAAAAGAAAATCTCTACTTAACCCAC.

V D I I E Q P F G E V E H A L S I R Q P F S F S F R D E L G  
Replicase 1a

TCGTGTTTTAGATCAATCTGATAATAATTGTTGGATTAGTACCACACTTATACAGTTGCAACTTACAAAGCTTTTGGATGATTCTATTGAG/  
AGCACAAAATCTAGTTAGACTATTATTAACAACCTAATCATGGTGTGAATATGTCAACGTTGAATGTTTCGAAAACCTACTAAGATAACTC

R V L D Q S D N N C W I S T T L I Q L O L T K L L D D S I E  
Replicase 1a

TGCAATTGTTTAAAGTTGGTAAAGTTGATTCAATTGTTCAAAAGTGTTATGAGTTGTCTCATTTAATTAGTGGTTCACCTGGTGATAGTGG/  
ACGTTAACAAATTTCAACCATTTCACCTAAGTTAACAAGTTTTCACAATACTCAACAGAGTAAATTAATCACCAAGTGAACCACTATCACC/

M Q L F K V G K V D S I V Q K C Y E L S H L I S G S L G D S G  
Replicase 1a

AAACTTCTTAGTGAACCTCTTAAAGATAAATATACATGTTCTATAACTTTTTGAGATGTCTTGTGATTGTGGTAAAAAGTTTGATGAGCAAG/  
TTTGAAGAATCACTGAAGAATTTCTATTATATGTACAAGATATTGAAAACCTCTACAGAACCTAACACCATTTTTCAAACCTACTCGTTCA

K L L S E L L K D K Y T C S I T F E M S C D C G K K F D E Q  
Replicase 1a

TGGTTGTTTGGTTTGGATTATGCCTTACACAAAACCTTTTTCAAAAAGGTGAGTGTGTTGTTTGTGCATAAAATGCAGACTTATAAGCTTGTTA  
ACCAACAAACAAAACCTAATACGGAATGTGTTTGGAAAAGTTTTTCCACTCACAACATAAACAGTATTTTACGTCTGAATATTGGAACAAT

G C L F W I M P Y T K L F Q K G E C C I - C H K M Q T Y K L V  
Replicase 1a

GTATGAAAGGTACTGGTGTGTTTGTACAGGATCCAGCACCTATTGACATTGATGCTTTCCCTGTTAGACCTATATGTTTCATCTGTATATTTA  
CATACTTTCCATGACCACACAAACATGTCCTAGGTCGTGGATAACTGTAACCTACGAAAGGGACAATCTGGATATACAAGTAGACATATAAAT

S M K G T G V F V Q D P A P I D I D A F P V R P I C S S V Y L  
Replicase 1a

GGTGTTAAGGGTTCTGGTCATTATCAAAACAAATTTATACAGTTTTGACAAAGCTATTGATGGTTTTGGTGTCTTTGACATTAAAAATAGTAG  
CCACAATTCCCAAGACCAGTAATAGTTTGTAAATATGTCAAACTGTTTCGATAACTACCAAAACCACAGAACTGTAATTTTTATCATC

G V K G S G H Y Q T N L Y S F D K A I D G F G V F D I K N S S  
Replicase 1a

IGTAAATACTGTTTGTGTTTGTGATGTTGATTTTCATAGTGTAGAAATAGAAGCTGGTGAAGTTAAACCTTTTGCTGTATATAAAAATGTTA  
ACAATTATGACAAACAAAACAACTACAACCTAAAAGTATCACATCTTTATCTTCGACCACTTCAATTTGGAAAACGACATATATTTTTACAAT

V N T V C F V D V D F H S V E I E A G E V K P F A V Y K N V  
Replicase 1a

AATTTTATTTAGGTGATATTTACACCTTGTAACCTGTGTTTCTTTGACTTTGTTGTCAATGCTGCTAATGAAAATCTCATGCATGGAGGC  
TTAAAATAAATCCACTATAAAGTGTGGAACATTTGACACAAAGAAAACCTGAAACAACAGTTACGACGATTACTTTTAGAGTACGTACCTCCG

K F Y L G D I S H L V N C V S F D F V V N A A N E N L M H G G  
Replicase 1a

GGTGTGCGCAGTGCTATTGATATTTTGAAGGTCAACTTCAGTCATTATCTAAAGATTACATTAGTAGTAATGGTCCACTTAAGGTTGG 42  
CCACAGCGTGCACGATAACTATAAACTGACTTCCAGTTGAAGTCAGTAATAGATTCTAATGTAATCATCATTACCAGGTGAATTCCAACC

G V A R A I D I L T E G Q L Q S L S K D Y I S S N G P L K V G  
Replicase 1a

AGCAGGTGTTATGTTGGAGTGTGAAAAATTCAATGTATTTAATGTTGTTGGTCCGCGAACTGGTAAACATGAGCATTTCATTACTTGTGTAAG 43  
TCGTCCACAATACAACCTCACACTTTTTAAGTTACATAAATTACAACAACCGGCGCTTGACCATTGTACTCGTAAGTAATGAACAACTTC

A G V M L E C E K F N V F N V V G P R T G K H E H S L L V E  
Replicase 1a

CTTATAATTCTATTTTATTTGAAAATGGTATTCCACTTATGCCTCTTCTTAGTTGTGGTATTTTGGTGTAAAGGATTGAAAATTCTCTTAAA 4  
GAATATTAAGATAAAATAAACTTTTACCATAAGGTGAATACGGAGAAGAATCAACACCATAAAAACCACATTCTCTAATTCTTAAGAGAATT

A Y N S I L F E N G I P L M P L L S C G I F G V R I E N S L K  
Replicase 1a

GCTTTGTTTAGTTGTGACATTAATAAACCATTGCAAGTTTTGTTTATTCTTCAAATGAAGAACAAGCTGTTCTTAAGTTTTAGATGGTTT 4  
CGAAACAAATCAACACTGTAATTATTTGGTAACGTTCAAAAACAAATAAGAAGTTACTTCTTGTTGACAAGAATTCAAAAATCTACCAA

A L F S C D I N K P L Q V F V Y S S N E E Q A V L K F L D G L  
Replicase 1a

AGATTTAACACCAGTCATTGACGATGTTGATGTTGTTAAACCTTTAGAGTTGAAGGTAATTTTTCATTCTTTGATTGTGGTGTCAATGCCT 4  
TCTAAATTGTGGTCAGTAAGTCTACAACACTACAACAATTGGAAAATCTCAACTTCCATTAAAAAGTAAGAACTAACACCACAGTTACGGA

D L T P V I D D V D V V K P F R V E G N F S F F D C G V N A  
Replicase 1a

TGGATGGTGATATTTACTTATTATTTACTAACTCTATTTTAATGTTGGATAAACAAGGACAATTATTGGACACAAAACCTTAATGGTATTTTG 4  
ACCTACCACTATAAATGAATAATAAATGATTGAGATAAAATTACAACCTATTTGTTCTGTTAATAACCTGTGTTTGAATTACCATAAAAC

L D G D I Y L L F T N S I L M L D K Q G Q L L D T K L N G I L  
Replicase 1a

CAACAGGCAGTTCTTGATTATCTTGCTACAGTTAAACTGTACCAGCTGGTAATTTGGTTAACTTGTGTTGAGAGTTGTACCATTATAT 4  
GTTGTCCGTCAAGAACTAATAGAACGATGTCAATTTTGACATGGTTCGACATTAAACCAATTTGAACAACAACCTCTCAACATGGTAAATATA

Q Q A V L D Y L A T V K T V P A G N L V K L V V E S C T I Y M  
Replicase 1a

GTGTGTTGTACCATCGATAAATGATCTTTCTTTTGATAAAAATCTTGGTCGTTGTGTGCGTAACTTAATAGATTGAAAACCTTGTGTTATTG 4  
CACACAACATGGTAGCTATTTACTAGAAAGAAAACCTATTTTAGAACAGCAACACACGCATTGGAATTATCTAATTTTGAACACAATAAC

C V V P S I N D L S F D K N L G R C V R K L N R L K T C V I  
Replicase 1a

CCAATGTTCTGCTATTGATGTTTTGAAAAAGCTTCTTCAAGTTTGACTTTAACTGTTAAATTTGTTGTAGAGAGTAATGTTATGGATGTT 4  
GGTTACAAGGACGATAACTACAAAACCTTTTGAAGAAAGTTCAAACCTGAAATTGACAATTTAAACAACATCTCTCATTACAATACCTACAA

A N V P A I D V L K K L L S S L T L T V K F V V E S N V M D V  
Replicase 1a

AACGACTGTTTTAAGAATGATAATGTAGTTTTGAAAATTACTGAAGATGGTATTAATGTTAAAGATGTTGTTGTTGAGTCTTCTAAGTCAC  
TTGCTGACAAAATTCTTACTATTACATCAAACTTTAATGACTTCTACCATAATTACAATTTCTACAACAACACTCAGAAGATTTCAGTG

N D C F K N D N V V L K I T E D G I N V K D V V V E S S K S  
Replicase 1a

TGGTAAACAATTGGGTGTTGTGAGTGATGGTGTGACTCTTTTGAAGGTGTTTTACCTATTAATACTGATACTGTCTTATCTGTAGCTCCA  
ACCATTGTTAACCACAACTCACTACCACAACCTGAGAAAACCTCCACAAAATGGATAATTATGACTATGACAGAATAGACATCGAGGT

G K Q L G V V S D G V D S F E G V L P I N T D T V L S V A P  
Replicase 1a

AAGTTGACTGGGTTGCTTTTTACGGTTTTGAAAAGGCAGCACTTTTTGCTTCTTTGGATGTAAAGCCATATGGTTACCCTAATGATTTTGT  
TTCAACTGACCCAACGAAAAATGCCAAAACCTTTCCGTCGTGAAAAACGAAGAAACCTACATTTCCGTATACCAATGGGATTACTAAAACA

E V D W V A F Y G F E K A A L F A S L D V K P Y G Y P N D F V  
Replicase 1a

GGTGGTTTTAGAGTTCTTGGGACCACCGACAATAATTGTTGGGTAAATGCAACTTGTATAATTTTACAGTATCTTAAGCCTACTTTTAAATC  
CCACCAAATCTCAAGAACCCTGGTGGCTGTTATTAACAACCCAATTACGTTGAACATATTAATGTCATAGAATTCGGATGAAAATTTAG

G G F R V L G T T D N N C W V N A T C I I L Q Y L K P T F K :  
Replicase 1a

TAAGGGTTTAAATGTTCTTTGGAACAAATTTGTTACAGGTGATGTTGGACCTTTTGTTAGTTTTATTTATTTTATAACTATGTCTTCAAAGG  
ATTCCCAAATTTACAAGAAACCTTGTTTAAACAATGTCCACTACAACCTGGAAAACAATCAAAATAAATAAATATTGATACAGAAGTTTCC

K G L N V L W N K F V T G D V G P F V S F I Y F I T M S S K  
Replicase 1a

GTCAAAGGGTGATGCTGAAGAGGCATTATCTAAATTGTCAGAGTATTTGATTAGTGATTCTATTGTTACTCTTGAACAATATTCAACTTGT  
CAGTTTTCCCACTACGACTTCTCCGTAATAGATTTAACAGTCTCATAAATAATCACTAAGATAACAATGAGAAGTTGTTATAAGTTGAACA

G Q K G D A E E A L S K L S E Y L I S D S I V T L E Q Y S T C  
Replicase 1a

GACATTTGTAAAAGTACTGTAGTTGAAGTTAAAAGTGCTGTTGTCTGTGCTAGTGTGCTTAAAGATGGTTGTGATGTTGGTTTTGTCCACA  
CTGTAAACATTTTCATGACATCAACTTCAATTTTCACGACAACAGACACGATCACACGAATTTCTACCAACACTACAACCAAAAACAGGTGT

D I C K S T V V E V K S A V V C A S V L K D G C D V G F C P H  
Replicase 1a

CAGACATAAATTGCGTTTCAGTGTAAAGTTTGTAAATGGACGTGTTGTTATTACCAATGTTGGTGAACCTATAATTTACACAACCTTCTAAGT  
GTCTGTATTTAACGCAAGTGCACAATTCAAACAATTACCTGCACAACAATAATGGTTACAACCACTTGGATATTAAGTGTGGAAGATTCA

R H K L R S R V K F V N G R V V I T N V G E P I I S Q P S K  
Replicase 1a

TGCTTAATGGTATTGCTTATAACAATTTTCAGGTTCTTTTGATAACGGTCACTATGTAGTTTATGATGCTGCTAATAATGCTGTCTATGAT  
ACGAATTACCATAACGAATATGTTGTAAAAGTCCAAGAAAACCTATTGCCAGTGATACATCAAATACTACGACGATTATTACGACAGATACTA

L L N G I A Y T T F S G S F D N G H Y V V Y D A A N N A V Y D  
Replicase 1a

GGTGCTCGTTTATTTGCTTCAGATTTGTCTACTTTAGCTGTTACAGCTATTGTTGTAGTAGGTGGTTGTGTAACATCTAATGTTCCACCAAT  
CCACGAGCAAATAAACGAAGTCTAAACAGATGAAATCGACAATGTCGATAACAACATCATCCACCAACACATTGTAGATTACAAGGTGGTTA  
G A R L F A S D L S T L A V T A I V V V G G C V T S N V P P I  
Replicase 1a

TGTTAGTGAGAAAATTTCTGTTATGGATAAACTTGATACTGGTGCACAAAAATTTTCCAATTTGGTGATTTTGTATGAATAACATTGTTCT  
ACAATCACTCTTTTAAAGACAATACCTATTGAACTATGACCACGTGTTTTTAAAAAGGTTAAACCACTAAAAACAATACTTATTGTAACAAG  
V S E K I S V M D K L D T G A Q K F F Q F G D F V M N N I V  
Replicase 1a

TGTTTTAACTTGGTTGCTTAGTATGTTTAGTCTTTTACGTACTTCTATTATGAAGCATGATATTAAGTTATTGCCAAGGCTCCTAAACGT  
ACAAAAATTGAACCAACGAATCATACAAATCAGAAAATGCATGAAGATAACTTCGTACTATAATTTCAATAACGGTTCCGAGGATTGCA  
L F L T W L L S M F S L L R T S I M K H D I K V I A K A P K R  
Replicase 1a

ACAGGTGTTATTTTGACACGTAGTTTTAAGTATAACATTAGATCTGCTTTGTTTGTGTAAGCAGAAGTGGTGTGTTATTGTTACTTTGTT  
TGTCACAATAAACTGTGCATCAAATTCATATTGTAATCTAGACGAAACAACAACATTTCTGCTTCACCACACAATAACAATGAAACAA  
T G V I L T R S F K Y N I R S A L F V V K Q K W C V I V T L F  
Replicase 1a

TAAGTTCTTATTGTTATTATATGCTATTTATGCACCTGTTTTATGATTGTGCAATTTAGTCCTTTAATAGTCTTTTATGTGGTGACATTG  
ATTCAAGAATAACAATAATATACGATAAATACGTGAACAAAAATACTAACACGTAAATCAGGAAAATTATCAGAAAATACACCACTGTAAC  
K F L L L L Y A I Y A L V F M I V Q F S P F N S L L C G D I  
Replicase 1a

TAAGTGGTTATGAAAAATCCACTTTTAATAAGGATATTTATTGTGGTAATTTCTATGGTTTGTAGATGTGTTTGTGTTAGTTATCAAGAGTTT  
ATTCACCAATACTTTTATGGTGAAAATTATTCCTATAAATAACACCATTAGATACCAAAACATTCTACACAAACAAATCAATAGTTCTCAA  
V S G Y E K S T F N K D I Y C G N S M V C K M C L F S Y Q E F  
Replicase 1a

AATGATTGGATCATACTAGTCTTGTGTTGGAAGCACATTCTGTATCCTATATTAATCAGTTTACAACCAATTTGTTATACTTGTATTTTGT  
TTACTAAACCTAGTATGATCAGAACAAACCTTCGTGTAAGCACTAGGATATAATTAGTCAAATGTTGGTAAACAATATGAACAATAAAACAA  
N D L D H T S L V W K H I R D P I L I S L Q P F V I L V I L L  
Replicase 1a

AATTTTGGTAATATGTATTTGCGTTTTGGACTTTTATATTTTGTGACAATTTATTAGTACTTTTGGTTCTTTCTTAGGCTTTTCATCAGA  
TTAAAAACCAATTATACATAAACGCAAAACCTGAAAATATAAAACAACGTGTTAAATAATCATGAAAACCAAGAAAGAATCCGAAAGTAGTCT  
I F G N M Y L R F G L L Y F V A Q F I S T F G S F L G F H Q  
Replicase 1a

AACAGTGGTTTTACATTTTGTCCGTTTGATGTTTTATGTAATGAGTTTTAGCTACATTTATTGTCTGCAAAATTTGTTTTATTTGTTAGA  
TTGTCACCAAAAAATGTAAACACGGCAAACTACAAAATACATTACTCAAAAATCGATGTAATAACAGACGTTTTAACAAAATAAACAAATCT  
K Q W F L H F V P F D V L C N E F L A T F I V C K I V L F V R  
Replicase 1a



CATATTATTGTTGGCTGTAATAATGCTGACTGTGTAGCTTGTCTCTAAAAGTGCTAGACTTAAACGTGTACCACTTCAAACATTATTAAT  
GTATAATAACAACCGACATTATTACGACTGACACATCGAACAAGATTTTCAGATCTGAATTTGCACATGGTGAAGTTTGATAATAATTAT  
H I I V G C N N A D C V A C S K S A R L K R V P L Q T I I N  
Replicase 1a

TATGCATAAATCATTCTATGTTAATGCTAATGGTGGTACTTGTCTCTGTAATAAACATAACTTCTTTTGTGTTAATTGTGATTCTTTTGGG  
ATACGTATTTAGTAAGATACAATTACGATTACCACCATGAACAAAGACATTATTTGTATTGAAGAAAACACAATTAACACTAAGAAAACCC  
M H K S F Y V N A N G G T C F C N K H N F F C V N C D S F G  
Replicase 1a

CTGGAATACTTTTATTAATGGTGATATTGCAAGAGAGCTTGGTAATGTTGTTAAAACAGCTGTTCAACCCACAGCTCCTGCATATGTTAT  
GACCATTATGAAAATAATTACCACTATAACGTTCTCTCGAACCATTACAACAATTTTGTGACAAGTTGGGTGTCGAGGACGTATACAATA  
P G N T F I N G D I A R E L G N V V K T A V Q P T A P A Y V I  
Replicase 1a

ATTGATAAGGTAGATTTTGTAAATGGATTTTATCGTCTTTATAGTGGTGACACTTTTTGGCGGTATGACTTTGACATTACTGAATCTAAGT  
TAACTATTCCATCTAAAACAATTACCTAAAATAGCAGAAATATCACCCTGTGAAAAACCGCCATACTGAAACTGTAATGACTTAGATTCA  
I D K V D F V N G F Y R L Y S G D T F W R Y D F D I T E S K  
Replicase 1a

TAGTTGTAAAGAGGTTCTGAAGAATTGTAATGTTTTAGAAAATTTTATTGTTTACAATAATAGTGGTAGTAACATTACACAGATTAATAAT  
ATCAACATTTCTCCAAGACTTCTTAACATTACAAAATCTTTTAAAATAACAAATGTTATTATCACCATCATTGTAATGTGTCTAATTTTTAC  
S C K E V L K N C N V L E N F I V Y N N S G S N I T Q I K N  
Replicase 1a

CTTGTTGTTATTTTTCTCAATTGTTGTGTGAACCTATAAAGTTGGTAAATTCAGAGTTGTTGTCAACTTTATCAGTTGATTTTAATGGTGT  
GAACACAAATAAAAAGAGTTAACAACACACTTGGATATTTCAACCATTAAAGTCTCAACAACAGTTGAAATAGTCAACTAAAATTACCACAA  
A C V Y F S Q L L C E P I K L V N S E L L S T L S V D F N G V  
Replicase 1a

TTGCATAAGGCATATGTTGATGTTTTGTGTAATAGTTTTTTAAGGAGCTAACTGCTAACATGTCCATGGCTGAATGTAAAGCTACACTTGG  
AACGTATTCGGTATACAACACAAAACACATTATCAAAAAAATTCCTCGATTGACGATTGTACAGGTACCGACTTACATTTTCGATGTGAACC  
L H K A Y V D V L C N S F F K E L T A N M S M A E C K A T L G  
Replicase 1a

TTTGACTGTTTCTGATGATGATTTTGTCTTCAGCTGTTGCCAATGCACATAGGTATGACGTTTGTCTTCAGATTTGTCAATTAATAATTTT  
AAACTGACAAAGACTACTACTAAAACAAAGTCGACAACGGTTACGTGTATCCATACTGCAAAACGAAAGTCTAAACAGTAAATTATTAATAA  
L T V S D D D F V S A V A N A H R Y D V L L S D L S F N N F  
Replicase 1a

TTATTTCTTATGCTAAACCTGAAGATAAGTTGTCCGTTTATGACATTGCTTGTGTATGCGTGCCGTTCTAAGGTTGTTAACCATAATGTT  
AATAAGAATACGATTTGGACTTCTATTCAACAGGCAATACTGTAACGAACAACATACGCACGGCCAAGATTCCAACAATTGGTATTACAA  
F I S Y A K P E D K L S V Y D I A C C M R A G S K V V N H N V  
Replicase 1a

TTAATCAAAGAGTCAATACCTATTGTTTGGGGTGTCAAGGACTTTAATACTCTTTCTCAAGAAGGTAAGAAGTACCTTGTTAAACAACCTAA  
AATTAGTTTCTCAGTTATGGATAACAAACCCACAGTTCCTGAAATTATGAGAAAGAGTTCTTCCATTCTTCATGGAACAATTTTGTGATT

L I K E S I P I V W G V K D F N T L S Q E G K K Y L V K T T K  
Replicase 1a

AGCAAAGGGTTTGACTTTTTTATTAACTTTAAATGATAACCAAGCAATTACACAAGTTCTGCTACTAGTATAGTTGCAAAACAGGGTGCTG  
TCGTTTCCCAAACGTAAAAAATAATTGAAAATTACTATTGGTTCGTTAATGTGTTCAAGGACGATGATCATATCAACGTTTGTCCCACGAC

A K G L T F L L T F N D N Q A I T Q V P A T S I V A K Q G A  
Replicase 1a

GTTTTAAACGTACTTATAATTTTCTGTGGTATGTATGTTTATTTGTTGTCATTGTTTATTGGTGTCTCATTTATTGATTATACAACCACT  
CAAAATTTGCATGAATATTAAGACACCATACATACAAATAAACAACAACGTAACAAATAACCACAGAGTAAATAACTAATATGTTGGTGA

G F K R T Y N F L W Y V C L F V V A L F I G V S F I D Y T T T  
Replicase 1a

GTAAGTAGCTTTTCATGGTTATGATTTTAAGTACATTGAGAATGGTCAGTTGAAGGTGTTTGAAGCACCTTTACACTGTGTTTCGTAATGTTTT  
CATTGATCGAAAGTACCAATACTAAAATTCATGTAACCTTTACCAGTCAACTTCCACAAACTTCGTGGAAATGTGACACAAGCATTACAAA

V T S F H G Y D F K Y I E N G Q L K V F E A P L H C V R N V F  
Replicase 1a

TGATAATTTTAATCAATGGCATGAGGCTAAGTTTGGTGTGTTACTACTAATAGTGATAAATGTCCTATAGTTGTTGGTGTTCAGAGCGTA  
ACTATTAAATTAGTTACCGTACTCCGATTCAAACCACAACAATGATGATTATCACTATTTACAGGATATCAACAACCACAAAGTCTCGCAT

D N F N Q W H E A K F G V V T T N S D K - C P I V V G V S E R  
Replicase 1a

TTAATGTTGTTCTGGTGTTCACAAATGTATATTTGGTAGGAAAGACTCTTGTTTTTACATTACAGGCTGCTTTTGGAAACACAGGTGTT  
AATTACAACAAGGACCACAAGGTTGTTTACATATAAACCATCCTTTCTGAGAACAAAAATGTAATGTCCGACGAAAACCTTTGTGTCCACAA

I N V V P G V P T N V Y L V G K T L V F T L Q A A F G N T G V  
Replicase 1a

TGTTATGACTTTGATGGTGTACCACTAGTGATAAGTGTATTTTAAATTCTGCTTGTACTAGGTTGGAAGGTTTGGGTGGTGACAATGTTTA  
ACAATACTGAAACTACCACAATGGTGATCACTATTACATAAAAAATTAAGACGAACATGATCCAACCTTCCAACCCACCACTGTTACAAAT

C Y D F D G V T T S D K C I F N S A C T R L E G L G G D N V Y  
Replicase 1a

TTGTTACAACACTGATCTTATTGAAGGTTCTAAACCTTATAGTATTTTACAGCCCAATGCTTATTATAAGTATGATGTTAAAAATTATGTAC  
AACAATGTTGTGACTAGAATAACTTCCAAGATTTGGAATATCATAAAATGTCGGGTACGAATAATATTCATACTACAATTTTAAATACATG

C Y N T D L I E G S K P Y S I L Q P N A Y Y K Y D V K N Y V  
Replicase 1a

GTTTTCCAGAAATTTAGCTAGAGGTTTGGCTTACGTACTATTAGAAGTTTGGCTACACGTTATTGTAGAGTTGGTGAATGCCGTGACTCA  
CAAAAGGTCTTTAAATCGATCTCCAAACCGAATGCATGATAATCTTGAAACCGATGTGCAATAACATCTCAACCACTTACGGCACTGAGT

R F P E I L A R G F G L R T I R T L A T R Y C R V G E C R D S  
Replicase 1a

CATAAAGGTGTTTGTGTTTGGTTTTGATAAATGGTATGTTAATGATGGACGTGTTGATGACGGTTACATTTGTGGTGATGGTCTTATAGACC  
GTATTTCCACAAACAAAACAAAACACTATTTACCATACAATTACTACCTGCACAACACTGCCAATGTAAACACCACTACCAGAATATCTGG

H K G V C F G F D K W Y V N D G R V D D G Y I C G D G L I D  
Replicase 1a

TCTTGTTAATGTACTCTCAATCTTTAGTTCATCTTTAGCGTTGTGGCTATGTCTGGACATATGTTGTTTAATTTCTTTTTGCAGCATT  
AGAACAATTACATGAGAGTTAGAAATCAAGTAGAAAATCGCAACACCGATACAGACCTGTATACAACAAATTTAAAGAAAAACGTCGTAAA

L V N V L S I F S S S F S V V A M S G H M L F N F L F A A F  
Replicase 1a

TTACATTTTTGTGCTTTTGTACTAAATTTAAACGTGTTTTGGTGATCTTCTTATGGTGTTTTACTGTTGTTGTGCAACTTTGAT  
AATGTAAAAACAGAAAAATCAATGATTTAAATTTGCACAAAAACCACTAGAAAGAATACCACAAAAATGACAACAAACACGTTGAAACTA

I T F L C F L V T K F K R V F G D L S Y G V F T V V C A T L I  
Replicase 1a

AATAACATTTCTTATGTTGTTACTCAAAATTTATTTTTATGTTGCTTTATGCTATTTTGTATTTTGTGTTTTACTAGGACAGTGCCTTATG  
TTATTGTAAAGAATACAACAATGAGTTTTAAATAAAAAATACAACGAAATACGATAAAACATAAAACAAAAATGATCCTGTCACGCAATAC

N N I S Y V V T Q N L F F M L L Y A I L Y F V F T R T V R Y  
Replicase 1a

TTGGATTTGGCATATTGCATACATTGTTGCATACTTCTTGTTAATACCATGGTGGCTTCTCACATGGTTTAGTTTTGCTGCATTTTTAGAG  
AACCTAAACCGTATAACGTATGTAACAACGTATGAAGAACAATTATGGTACCACCGAAGAGTGACCAAATCAAACGACGTAAAAATCTC

W I W H I A Y I V A Y F L L I P W W L L T W F S F A A F L E  
Replicase 1a

TTTTACCTAATGTTTTAAGTTAAAAATCTCTACTCAATTGTTTGAAGGTGATAAGTTTATAGGTACTTTTGAGAGTGCTGCTGCAGGTACA  
AAAATGGATTACAAAAATTCAATTTTTAGAGATGAGTTAACAACTTCCACTATTCAAATATCCATGAAACTCTCACGACGACGTCCATGT

L L P N V F K L K I S T Q L F E G D K F I G T F E S A A A G T  
Replicase 1a

TTTGTCTTGACATGCGTTCTTATGAAAGGCTGATAAATACTATTTACCTGAGAACTTAAGAATTATGCTGCAAGTTATAATAAATATAA  
AAACAAGAACTGTACGCAAGAATACTTCCGACTATTTATGATAAAGTGGACTCTTTGAATTCTTAATACGACGTTCAATATTATTATATT

F V L D M R S Y E R L I N T I S P E K L K N Y A A S Y N K Y  
Replicase 1a

ATATTATAGTGGTAGTGTCTAGTGAGGCTGATTATCGITGTGCTTGTATGCTCATTTAGCCAAGGCTATGTTAGATTAGGCAAAAGATCATA  
TATAATATCACCATCAGATCACTCCGACTAATAGCAACACGAACAATACGAGTAAATCGGTTCCGATACAATCTAATGCGTTTTCTAGTAT

Y Y S G S A S E A D Y R C A C Y A H L A K A M L D Y A K D H  
Replicase 1a

ATGACATGTTATATTCTCCACCTACCATTAGCTACAATTCCACCTTACAATCTGGTCTTAAGAAGATGGCACAACCATCTGGTTGTGTTGAG  
TACTGTACAATATAAGAGGTGGATGGTAATCGATGTTAAGGTGGAATGTTAGACCAGAATTCTTCTACCGTGTGGTAGACCAACACAACCTC

N D M L Y S P P T I S Y N S T L O S G L K K M A Q P S G C V E  
Replicase 1a

AGATGTGTGGTTTCGCGTCTGTTATGGTAGTACTGTGCTTAATGGAGTTTGGTTAGGTGACACTGTTACTTGTCTAGACATGTCATAGCACC  
TCTACACACCAAGCGCAGACAATACCATCATGACACGAATTACCTCAAACCAATCCACTGTGACAATGAACAGGATCTGTACAGTATCGTGG

R C V V R V C Y G S T V L N G V W L G D T V T C P R H V I A P  
Replicase 1a

ATCAACCACTGTTCTTATTGATTATGATCATGCATATAGTACTATGCGTTTGATAATTTTTTCAGTGTCTCATAATGGTGTCTTCTTGGGAG  
TAGTTGGTGACAAGAATAACTAATACTAGTACGTATATCATGATACGCAAACGTATTAAGTACAGAGTATTACCACAGAAGAACCCTC

S T T V L I D Y D H A Y S T M R L H N F S V S H N G V F L G  
Replicase 1a

TTGTTGGTGTTACAATGCATGGTTCTGTGTTGCGTATTAAGGTTTCACAATCTAATGTACATACCTAAACATGTTTTTAAACGTTGAAA  
AACAACCACAATGTTACGTACCAAGACACAACGCATAATTCCAAAGTGTTAGATTACATGTATGTGGATTTGTACAAAAATTTGCAACTTT

V V G V T M H G S V L R I K V S Q S N V H T P K H V F K T L K  
Replicase 1a

CCTGGTGCTTCTTTAATATTTTAGCATGTTATGAAGGTATTGCATCTGGTGTGTTTGGTGTTAATTTACGTACAACTTTACTATTAAAGG  
GGACCACGAAGAAAATTATAAAATCGTACAATACTTCCATAACGTAGACCACAAAAACCACAATTAATGCATGTTTGAATGATAATTTCC

P G A S F N I L A C Y E G I A S G V F G V N L R T N F T I K G  
Replicase 1a

TTCTTTTATAAATGGAGCTTGTGGTTCTCCTGGTTATAATGTTAGAAATGATGGTACTGTTGAGTTTTGTTATTTACACCAAATTGAGTTAG  
AAGAAAATATTTACCTCGAACACCAAGAGGACCAATATTACAATCTTTACTACCATGACAACCTCAAAACAATAAATGTGGTTTAACTCAATC

S F I N G A C G S P G Y N V R N D G T V E F C Y L H Q I E L  
Replicase 1a

GTAGTGGTGCTCATGTTGGTTCTGATTTTACTGGTAGTGTGTTATGGTAATTTGATGACCAACCTAGTTTGCAAGTTGAGAGTGCCAACCTT  
CATCACCACGAGTACAACCAAGACTAAAATGACCATCACAATACCATTAAACTACTGGTTGGATCAAACGTTCAACTCTCAGGTTGGAA

G S G A H V G S D F T G S V Y G N F D D Q P S L Q V E S A N L  
Replicase 1a

ATGCTATCAGATAATGTTGTTGCCTTTTTGTATGCTGCTTTGTTGAATGGTTGTAGGTGGTGGTTGCGTTCAACTAGAGTTAATGTTGATGG  
TAGGATAGTCTATTACAACAACGGAAAAACATACGACGAAACAACTTACCAACATCCACCACCAACGCAAGTTGATCTCAATTACAACCTACC

M L S D N V V A F L Y A A L L N G C R W W L R S T R V N V D G  
Replicase 1a

TTTTAATGAATGGGCTATGGCTAATGGTTATACAATTGTTTCTAGTGTGAGTGCTATTCTATTTTGGCAGCAAAAACCTGGTGTAGTGTG  
AAAATTACTTACCGGATACCGATTACCAATATGTTAACAAGATCACAACCTCAGGATAAGATAAAACCGTCGTTTTTGACCACAATCACAAC

F N E W A M A N G Y T I V S S V E C Y S I L A A K T G V S V  
Replicase 1a

AACAATTGTTAGCTTCCATTCAACATCTTCATGAAGGTTTTGGTGGTAAAAACATACTTGGTTATTCTAGTTTATGTGATGAGTTCACACTA  
TTGTTAACAATCGAAGGTAAGTTGTAGAAGTACTCCAAAACCACTTTTTGTATGAACCAATAAGATCAAATACACTACTCAAGTGTGAT

E Q L L A S I Q H L H E G F G G K N I L G Y S S L C D E F T L  
Replicase 1a

13/87

GCTGAAGTTGTGAAGCAGATGTATGGTGTAACTTGCAAAGTGGTAAGGTTATTTTTGGTTTAAAAACAATGTTTTATTTAGCGTTTTCT  
CGACTTCAACACTTCGTCTACATACCACAATTGAACGTTTCACCATTCGAATAAAAAACCAATTTTTGTTACAAAAATAAATCGCAAAAGA.

A E V V K Q M Y G V N L Q S G K V I F G L K T M F L F S V F  
Replicase 1a

CACAATGTTTTGGGCAGAACTCTTTATTATACAAACACTATATGGATAAACCTGTTATACTTACACCTATATTTTGTCTTACTTTTGT  
GTGTTACAAAACCGTCTTGAGAAATAAATATGTTTGTGATATACCTATTTGGGACAATATGAATGTGGATATAAAACAAATGAAAACAAA

T M F W A E L F I Y T N T I W I N P V I L T P I F C L L L F  
Replicase 1a

TGTCATTAGTTTTAACTATGTTTCTTAAACATAAGTTTTTGTCTTGAAGTATTTTTATTACCTACTGTTATTGCAACTGCTTTATATAA  
ACAGTAATCAAAATTGATACAAAGAATTTGTATTCAAAAACAAAACGTTTCATAAAAATAATGGATGACAATAACGTTGACGAAATATATT

L S L V L T M F L K H K F L F L Q V F L L P T V I A T A L Y N  
Replicase 1a

TGTGTTTTGGATTATTACATAGTAAAATTTTTGGCTGACCATTTAACTATAATGTTTCAGTATTACAAATGGATGTTTCAGGGTTTAGTTA  
ACACAAAACCTAATAATGTATCATTTTTAAAAACCGACTGGTAAAATTGATATTACAAAGTCATAATGTTTACCTACAAGTCCCAAATCAATT

C V L D Y Y I V K F L A D H F N Y N V S V L Q M D V Q G L V I  
Replicase 1a

TGTTTTGGTCTGTTTATTTGTTGTATTTTTACACACATGGCGTTTTCTAAAGAACGTTTCACACATTGGTTTACATATGTGTGTTCTCTTA  
ACAAAACCGACAAATAAACACATAAAAAATGTGTGTACCGCAAAAAGATTTCTTGCAAAGTGTGTAACCAATGTATACACACAAGAGAAT

V L V C L F V V F L H T W R F S K E R F T H W F T Y V C S L  
Replicase 1a

TAGCAGTTGCTTACACTTATTTTTATAGTGGTGACTTTTTGAGTTTGCTTGTTATGTTTTATGTGCTATATCTAGTGATTGGTACATTGGT  
ATCGTCAACGAATGTGAATAAAAAATATCACCCTGAAAACTCAAACGAACAATACAAAAATACACGATATAGTCACTAACCATGTAACCA

I A V A Y T Y F Y S G D F L S L L V M F L C A I S S D W Y I G  
Replicase 1a

GCCATTGTTTTAGGTTGTCACGTTTGATTATTTTTTTTACCTGAAAGTGTATTTAGTGTTTTTGGTGATGTGAAACTCACTTTAGTTGT  
CGGTAACAAAAATCCAACAGTGCAAACTAATATAAAAAAAGTGGACTTTACATAAATCACAAAAACCACTACACTTTGAGTGAAATCAACA

A I V F R L S R L I I F F S P E S V F S V F G D V K L T L V V  
Replicase 1a

LIATITAAATIGTGTTATTTAGITTTGTACTTATTTGGGGCATTTTGTATTGGTTCAATAGGTTTTTAAATGTACTATGGGTGTTTATGATT  
AATAAATTAACACCAATAAATCAAACATGAATAACCCCGTAAACATAACCAAGTTATCCAAAAATTTACATGATACCCACAAATACTAA

Y L I C G Y L V C T Y W G I L Y W F N R F F K C T M G V Y D  
Replicase 1a

TTAAGGTGAGTGCTGCTGAATTTAAATACATGGTTGCTAATGGACTTCATGCACCATATGGACCTTTTGATGCACTTTGGTTATCATTCAAA  
AATCCACTCAGCAGACTTAAATTTATGTACCAACGATTACCTGAAGTACGTGGTATACCTGGAAAACTACGTGAAACCAATAGTAAGTTT

F K V S A A E F K Y M V A N G L H A P Y G P F D A L W L S F K  
Replicase 1a

TTACTTGGTATTGGTGGTGACCGTTGTATAAAAAATTTCAACTGTCCAATCCAACTGACTGATTGGAAGTGTACTAATGTTGTGTTATTGGG  
AATGAACCATAACCACCACTGGCAACATATTTTTAAAGTTGACAGGTTAGGTTTGACTGACTAACTTCACATGATTACAACACAATAACCC

L L G I G G D R C I K I S T V Q S K L T D L K C T N V V L L G  
Replicase 1a

TTGTTTGTCTAGTATGAACATTGCAGCTAATTCTAGTGAATGGGCTTATTGTGTTGATTTACACAATAAGATTAATCTTTGTGATGACCCAG  
AACAAACAGATCATACTTGTAAAGTCGATTAAGATCACTTACCCGAATAACACAATAAATGTGTTATTCTAATTAGAAACACTACTGGGTC

C L S S M N I A A N S S E W A Y C V D L H N K I N L C D D P  
Replicase 1a

AAAAAGCTCAAGGTATGTTGTTAGCACTCCTTGCCTTCTTTCTAAGTAAACATAGTGATTTTGGTCTTGATGGCCTTATTGATTCTTATTTT  
TTTTTCGAGTTCATACAACAATCGTGAGGAACGCAAGAAAGATTCAATTTGTATCACTAAAACCAAGTACCGGAATAACTAAGAATAAAA

E K A Q G M L L A L L A F F L S K H S D F G L D G L I D S Y F  
Replicase 1a

GATAATAGTAGCACCTGCAGAGTGTGCTTCATCATTGTTAGTATGCCATCATATATTGCTTATGAAAATGCTAGACAAGCTTATGAGGA  
CTATTATCATCGTGGGACGTCTCACAACGAAGTAGTAAACAATCATACGGTAGTATATAACGAATACTTTACGATCTGTTTGAATACTCCT

D N S S T L Q S V A S S F V S M P S Y I A Y E N A R Q A Y E D  
Replicase 1a

TGCTATTGCTAATGGATCTTCTTCTCAACTTATTAACAATTGAAGCGTGCCATGAATATCGCAAAGTCTGAATTTGATCATGAGATATCTG  
ACGATAACGATTACCTAGAAGAAGAGTTGAATAATTTGTTAACTTCGCACGGTACTTATAGCGTTTCAGACTTAACTAGTACTCTATAGAC

A I A N G S S S Q L I K Q L K R A M N I A K S E F D H E I S  
Replicase 1a

TTCAGAAGAAAATTAATAGAATGGCTGAACAAGCTGCTACTCAGATGTATAAAGAAGCACGCTCTGTTAATAGAAAATCTAAAGTTATTAGT  
AAGTCTTCTTTAATTATCTTACCGACTTGTTGACGATGAGTCTACATATTTCTTCGTGCGAGACAATTATCTTTTAGATTTCATAATCA

V Q K K I N R M A E Q A A T Q M Y K E A R S V N R K S K V I S  
Replicase 1a

GCTATGCACTCTTTACTTTTTGGAATGTTAAGACGTTTGGATATGTCTAGTGTGAACTGTTTTGAATTTAGCACGTGATGGTGTGTTGCC  
CGATACGTGAGAAATGAAAAACCTTACAATTCTGCAACCTATACAGATCACAACCTTGACAAAACCTTAAATCGTGCACTACCACAACACGG

A M H S L L F G M L R R L D M S S V E T V L N L A R D G V V P  
Replicase 1a

ATTGTCAGTTATACCTGCAACTTCAGCTTCCAACTAACTATTGTTAGTCCAGATCTTGAATCTTATTCTAAGATTGTTTGTGATGGTTCTG  
TAACAGTCAATATGGACGTTGAAGTCGAAGGTTTGATTGATAACAATCAGGTCTAGAAGTTAGAATAAGATTCTAACAAACACTACCAAGAC

L S V I P A T S A S K L T I V S P D L E S Y S K I V C D G S  
Replicase 1a

TTCATTATGCTGGAGTTGTTTGGACACTTAATGATGTTAAAGACAATGATGGTAGACCTGTTTCATGTTAAAGAGATTACAAGGGAGAATGTT  
AAGTAATACGACCTCAACAAACCTGTGAATTACTACAATTTCTGTTACTACCATCTGGACAAGTACAATTTCTCTAATGTTCCCTCTTACAA

V H Y A G V V W T L N D V K D N D G R P V H V K E I T R E N V  
Replicase 1a

GAAACTTTGACATGGCCTCTTATCCTTAATTGTGAACGTGTTGTTAACTTCAAAATAATGAAATTATGCCTGGTAACTTAAGCAAAAA  
CTTTGAAACTGTACCGGAGAATAGGAATTAACACTTGCACAACAATTTGAAGTTTATTACTTTAATACGGACCATTGAATTGTTTTT  
E T L T W P L I L N C E R V V K L Q N N E I M P G K L K Q K  
Replicase 1a

TATGAAAGCTGAGGGTGATGGTGGTGTGTTAGGTGATGGTAATGCTTTGTATAATACTGAGGGTGGTAAACTTTTATGTATGCTTATAT  
ATACTTTGACTCCCCTACCACCACAAAATCCACTACCATTACGAAACATATTATGACTCCCACCATTGAAAATACATACGAATATA  
M K A E G D G G V L G D G N A L Y N T E G G K T F M Y A Y I  
Replicase 1a

CTAATAAGCTGACCTTAAATTTGTTAAGTGGGAGTATGAGGGTGGTTGCAACACAATCGAGTTAGACTCTCCTTGTCGATTTATGGTCGA  
GATTATTTGACTGGAATTTAAACAATTCACCTCATACTCCACCAACGTTGTGTTAGCTCAATCTGAGAGGAACAGCTAAATACCAGCT  
S N K A D L K F V K W E Y E G G C N T I E L D S P C R F M V I  
Replicase 1a

ACACCTAATGGTCCTCAAGTGAAGTATTTGTATTTTGTAAAAATTTAAATACCTTACGTAGAGGTGCCGTTCTTGGTTTTATAGGTGCCA  
TGTGGATTACCAGGAGTTCACCTTCATAAACATAAAACAATTTTAAATTTATGGAATGCATCTCCACGGCAAGAACCAAAATATCCACGGT  
T P N G P Q V K Y L Y F V K N L N T L R R G A V L G F I G A  
Replicase 1a

AATTCGTCTACAAGCTGGTAAACAACTGAATTGGCTGTTAATTCTGGACTTTTAACTGCTTGTGCTTTTTCTGTTGATCCAGCAACCACT  
TTAAGCAGATGTTGACCATTTGTTTGACTTAACCGACAATTAAGACCTGAAAATTGACGAACACGAAAAAGACAAGTAGGTGCTTGGTGA  
I R L Q A G K Q T E L A V N S G L L T A C A F S V D P A T T  
Replicase 1a

ACTTGAAGCTGTTAAACATGGTGCAAAACCTGTAAGTAATTGTATTAAGATGTTATCTAATGGTGTGTAATGGTCAAGCTATAACAAC  
TGAACCTTCGACAATTTGTACCAGTTTTGGACATTCATTAACATAATTCTACAATAGATTACCACGACCATTACCAGTTCGATATTGTTG  
Y L E A V K H G A K P V S N C I K M L S N G A G N G Q A I T T  
Replicase 1a

AGTGTAGATGCTAACACCAATCAAGATTCTTATGGTGGAGCGTCTATTTGTTTGTATTGTGGGGCCACGTTCTCACCCTAGTATGGATGC  
TCACATCTACGATTGTGGTTAGTTCTAAGAATACCACCTCGCAGATAAACAAACATAACAGCCCGGTGCAAGGAGTGGGATCATACCTACC  
S V D A N T N Q D S Y G G A S I C L Y C R A H V P H P S M D I  
Replicase 1a

TTACTGTAAGTTTAAGGGTAAATGTGTTTCAGGTTCTATTGGTTGTTGGATCCTATTAGGTTTGTGTTAGAAAAATATGTGTGTAATGTTT  
AATGACATTCAAATTCCTTTACACAAGTCCAAGGATAACCAACAAACCTAGGATAATCCAAAACAAATCTTTTATTACACACATTACAAA  
Y C K F K G K C V Q V P I G C L D P I R F C L E N N V C N V  
Replicase 1a

GTGGTTGTTGGTTGGGACACGGGTGTGCTTGTGATCGTACAACCATTCAAAGTGTGACATTTCTTATTTAAACGAGCAAGGGGTTCTAGTG  
CACCAACAACCAACCCTGTGCCACACGAACACTAGCATGTTGGTAAGTTTACAACGTAAAGAATAAATTTGCTCGTTCCCCAAGATCAC  
C G C W L G H G C A C D R T T I Q S V D I S Y L N E Q G V L V  
Replicase 1a

R A R G S S  
Replicase 1b

CAGCTCGACTAGAACCCTGTAATGGCACGGACATCGATAAGTGTGTTCTGCTTTTGACATTTATAATAAAAAATGTTTCATTCTTGGGTAAG  
GTCGAGCTGATCTTGGGACATTACCGTGCCGTGTAGCTATTACACAAGCACGAAACTGTAAATATTATTTTACAAAGTAAGAACCCATTC

Q L D  
Replicase 1a

A A R L E P C N G T D I D K C V R A F D I Y N K N V S F L G K  
Replicase 1b

TGTTTGAAGATGAACGTGTTCGTTTTAAAAATGCTGATCTTAAGGATGGTTATTTTGTATAAAGAGGTGTAAGTCGGTTATGGAACA  
ACAAACTTCTACTTGACACAAGCAAAATTTTACGACTAGAATTCCTACCAATAAAACAATATTTCTCCACATGATTCAGCCAATACCTTGT

C L K M N C V R F K N A D L K D G Y F V I K R C T K S V M E H  
Replicase 1b

CGAGCAATCCATGTATAACCTACTTAACTTTTCTGGTGCTTTGGCTGAGCATGATTTCTTTACTTGGAAGATGGCAGAGTCATTTATGGTA  
GCTCGTTAGGTACATATTGGATGAATTGAAAAGACCACGAAACCGACTCGTACTAAAGAAATGAACCTTTCTACCGTCTCAGTAAATACCAT

E Q S M Y N L L N F S G A L A E H D F F T W K D G R V I Y G  
Replicase 1b

ATGTTAGTAGACATAATCTTACTAAATATACTATGATGGACTTGGTTTATGCTATGCGTAACTTTGATGAACAAAATTGTGATGTTCTAAAA  
TACAATCATCTGTATTAGAATGATTTATATGATACTACCTGAACCAATACGATACGCATTGAACTACTTGTTTAACTACAAGATTTT

N V S R H N L T K Y T M M D L V Y A M R N F D E Q N C D V L K  
Replicase 1b

GAAGTATTAGTTTTAACTGGTTGTTGTGACAATTCTTATTTTGATAGTAAGGGTTGGTATGACCCAGTTGAAAATGAAGATATACATAGAGT  
CTTCATAATCAAAATTGACCAACAACACTGTTAAGAATAAACTATCATTCCCAACCATACTGGGTCAACTTTTACTTCTATATGTATCTCA

E V L V L T G C C D N S Y F D S K G W Y D P V E N E D I H R V  
Replicase 1b

TTATGCATCTCTTGGCAAAATTGTAGCTAGAGCTATGCTTAAATGCGTTGCTCTATGTGATGCGATGGTTGCTAAAGGTGTTGTTGGTGT  
AATACGTAGAGAACCGTTTTAACATCGATCTCGATACGAATTTACGCAACGAGATACACTACGCTACCAACGATTTCCACAACAACCACAAA

Y A S L G K I V A R A M L K C V A L C D A M V A K G V V G V  
Replicase 1b

TAACATTAGATAACCAAGATCTTAATGGTAACTTTTATGATTTTGGTGATTTTGTGTTAGCTTACCTAATATGGGTGTTCCCTGTTGTACA  
ATTGTAATCTATTGGTTCTAGAATTACCATGAAAATACTAAACCCTAAACAACAATCGAATGGATTATACCCACAAGGGACAACATGT

L T L D N Q D L N G N F Y D F G D F V V S L P N M G V P C C T  
Replicase 1b



17/87

TCATATTATTCTTATATGATGCCTATTATGGGTTTAACTAATTGTTTAGCTAGTGAGTGTTTTGTCAAGAGTGATATTTTTGGTAGTGAT  
AGTATAATAAGAATATACTACGGATAATACCCAAATTGATTAACAAATCGATCACTCACAAAACAGTTCTCACTATAAAAACCATCACTA  
S Y Y S Y M M P I M G L T N C L A S E C F V K S D I F G S D  
Replicase 1b

TAAACTTTTGATTGCTTAAGTATGATTTCACTGAACATAAAGAAAATTTATTCAATAAGTACTTTAAGCATTGGAGTTTTGATTATCAT  
ATTTTGAAAACATAACGAATTCATACTAAAGTGACTTGTATTTCTTTTAAATAAGTTATTCATGAAATTCGTAACCTCAAACTAATAGTA  
K T F D L L K Y D F T E H K E N L F N K Y F K H W S F D Y H  
Replicase 1b

CTAATTGTAGTGACTGTTATGATGATATGTGTGTTATACATTGTGCTAATTTAATACACTATTGCCCACAACTATACCAGGTACTGCTTT  
GATTAACATCACTGACAATACTACTATACACACAATATGTAACACGATTAAATTTATGTGATAAACGGTGTTGATATGGTCCATGACGAAA  
P N C S D C Y D D M C V I H C A N F N T L F A T T I P G T A F  
Replicase 1b

GGTCCACTATGTCGTAAAGTTTTATAGATGGTGTTCCTACTTGTACAACCTGCTGGTTATCATTTTAAGCAATTAGGTTTGGTTTGAATA  
CCAGGTGATACAGCATTTCAAAAATATCTACCACAAGGTGAACAATGTTGACGACCAATAGTAAAATTCGTTAATCCAAACCAACCTTAT  
G P L C R K V F I D G V P L V T T A G Y H F K Q L G L V W N  
Replicase 1b

AGATGTTAACACACACTCAGTTAGGTTGACAATCACTGAACCTTTGCAATTTGTTACTGACCTTCCTTGATAATAGCTTCTTCTCCAGCA  
TCTACAATTGTGTGTGAGTCAATCCAACCTGTTAGTGACTTGAAAACGTTAAACAATGACTGGGAAGGAAGTATTATCGAAGAAGAGGTCGT  
D V N T H S V R L T I T E L L Q F V T D P S L I I A S S P A  
Replicase 1b

TCGTTGATCAACGCACTATTTGTTTTCTGTTGCAGCATTGAGTACTGGTTTGACAAATCAAGTTGTTAAGCCAGGTCATTTTAATGAAGA  
AGCAACTAGTTGCGTGATAAACAAGACAACGTCGTAACCTCATGACCAAACTGTTTAGTTCAACAATTCGGTCCAGTAAATTACTTCTI  
L V D Q R T I C F S V A A L S T G L T N Q V V K P G H F N E E  
Replicase 1b

TTTTATACTTTCTTCGTTTAAGAGGTTTCTTTGATGAAGTTCTGAACTTACATTAAACATTTCTTCTTCGCACAGAATGGTGATGCTG  
AAAATATTGAAAGAAGCAATTCTCAAAGAACTACTTCCAAGACTTGAATGTAATTTTGTAAGAAGAAGCGTGTCTTACCACTACGAC  
F Y N F L R L R G F F D E G S E L T L K H F F F A Q N G D A  
Replicase 1b

IGTTAAAGATTTTGACITTTACCGTTATAAIAAGCCTACCAITTTAGATATTTGTCAAGCTAGAGTTACATATAAGATAGTCTCTCGTTAT  
ACAATTTCTAAAACGAAAATGGCAATATTATTCGGATGGTAAAATCTATAAACAGTTTCGATCTCAATGTATATTCTATCAGAGAGCAATA  
V K D F D F Y R Y N K P T I L D I C O A R V T Y K I V S R Y  
Replicase 1b

TTGACATTTATGAAGGTGGCTGTATTAAGGCATGTGAAGTTGTTGTAACAAATCTTAATAAGAGTGCTGGTTGGCCATTAAATAAGTTTGGT  
AACTGTAAATACTTCCACCGACATAATTCGTACACTTCAACAACATTGTTTAGAATTATTCTCAGACCAACCGGTAATTTATTCAAACCA  
F D I Y E G G C I K A C E V V V T N L N K S A G W P L N K F G  
Replicase 1b

AAAGCTAGTTTGTATTACGAATCTATATCTTATGAAGAACAGGATGCTTTGTTTGCTTTGACAAAGCGTAATGTCCTCCCTACTATGACACA  
TTTCGATCAAACATAATGCTTAGATATAGAATACTTCTTGTCTACGAAACAAACGAACTGTTTCGCATTACAGGAGGGATGATACTGTGT

K A S L Y Y E S I S Y E E Q D A L F A L T K R N V L P T M T Q  
Replicase 1b

GCTGAATCTTAAGTATGCTATTAGTGGTAAAGAACGTGCTAGAACTGTTGGTGGTGTCTCTGTTGTCCACAATGACCACAAGACAATACC  
CGACTTAGAATTCATACGATAATCACCATTCTTGCACGATCTTGACAACCACCACAAAGAGACAACAGGTGTTACTGGTGTCTGTATGG

L N L K Y A I S G K E R A R T V G G V S L L S T M T T R Q Y  
Replicase 1b

ATCAAAAACATCTTAAATCCATTGTTAATACACGCAATGCCACTGTTGTTATTGGTACTACCAAATTTTATGGTGGTTGGAATAATATGTTG  
TAGTTTTGTAGAATTTAGGTAACAATTATGTGCGTTACGGTGACAACAATAACCATGATGGTTTAAAATACCACCAACCTTATTATACAAC

H Q K H L K S I V N T R N A T V V I G T T K F Y G G W N N M L  
Replicase 1b

CGTACTTTAATTGATGGTGTGAAAACCCTATGCTCATGGGTTGGGATTATCCCAAATGTGATAGAGCTTTCCTAACATGATACGTATGAT  
GCATGAAATTAAC TACCACAAC TTTTGGGATACGAGTACCCAACCCTAATAGGGTTTACACTATCTCGAAACGGATTGTACTATGCATACTA

R T L I D G V E N P M L M G W D Y P K C D R A L P N M I R M I  
Replicase 1b

TTCAGCCATGGTGTGGGTTCTAAGCATGTTAATTGTTGTACTGTAACAGATAGGTTTTATAGGCTTGGTAACGAGTTGGCACAAGTTTTAA  
AAGTCGGTACCACAACCAAGATTTCGTACAATTAACAACATGACATTGTCTATCCAAAATATCCGAACCATTGCTCAACCGTGTTCAAAATT

S A M V L G S K H V N C C T V T D R F Y R L G N E L A Q V L  
Replicase 1b

CAGAAGTTGTTTATTCTAATGGTGGTTTTATTTAAGCCAGGTGGTACGACTTCTGGTGACGCTAGTACAGCTTATGCTAATTCTATTTTT  
GTCTTCAACAAATAAGATTACCACCAAAAATAAAATTCGGTCCACCATGCTGAAGACCACTGCGATCATGTGCAATACGATTAAGATAAAAA

T E V V Y S N G G F Y F K P G G T T S G D A S T A Y A N S I F  
Replicase 1b

AACATTTTTCAAGCCGTGAGTTCTAACATTAACAGGTTGCTTAGTGTCCTATCAGATTCATGTAATAATGTTAATGTTAGGGATCTACAACG  
TTGTAAAAAGTTCGGCACTCAAGATTGTAATTGTCCAACGAATCACAGGGTAGTCTAAGTACATTATTACAATTACAATCCCTAGATGTTGC

N I F Q A V S S N I N R L L S V P S D S C N N V N V R D L Q R  
Replicase 1b

ACGTCTGTATGATAATTGCTATAGGTTAACTAGTGTGAAGAGTCATTGATGATTATTATGGTTATCTTAGGAAACATTTTTCAATGA  
TGCAGACATACTATTAACGATATCCAATTGATCACAACCTCTCAGTAAGTAACTACTAATAATACCAATAGAATCCTTTGTA AAAAGT TACT

R L Y D N C Y R L T S V E E S F I D D Y Y G Y L R K H F S M  
Replicase 1b

TGATTCTCTCTGATGACGGTGTGTCTGTTATAACAAGGATTATGCTGAGTTAGGTTATATAGCAGACATTAGTGCTTTTAAAGCCACTTTG  
ACTAAGAGAGACTACTGCCACAACAGACAATATTGTTCTAATACGACTCAATCCAATATATCGTCTGTAATCACGAAAATTTTCGGTGAAC

M I L S D D G V V C Y N K D Y A E L G Y I A D I S A F K A T L  
Replicase 1b

TATTACCAGAATAATGCTTTTATGAGTACTTCTAAATGTTGGGTTGAAGAAGATTAACTAAGGGACCACATGAGTTTTGTTCCAGCAT/  
ATAATGGTCTTATTACAGAAATACTCATGAAGATTTACAACCCAACCTCTTCTAAATTGATTCCCTGGTGTACTCAAAACAAGGGTCGTA1  
Y Y Q N N V F M S T S K C W V E E D L T K G P H E F C S Q H  
Replicase 1b

TATGCAAATAGTTGATAAAGATGGTACCTATTATTTGCCTTACCCAGATCCTAGTAGGATCTTGTGAGCTGGTGTGTTTTGTTGATGATGTT  
ATACGTTTATCAACTATTTCTACCATGGATAATAAACGGAATGGGTCTAGGATCATCCTAGAACAGTCGACCACAAAAACAATACTACTACAA  
M Q I V D K D G T Y Y L P Y P D P S R I L S A G V F V D D V  
Replicase 1b

TTAAGACAGATGCTGTTGTTTTGTTAKAACGTTATGTGTCTTTAGCTATTGATGCATACCCCTCTTTCAAAACACCCTAATTCTGAATATCG  
AATTCTGTCTACGACAACAAACAATMTTGCAATACACAGAAATCGATAACTACGTATGGGAGAAAGTTTTGTGGGATTAAGACTTATAGC  
V K T D A V V L L ? R Y V S L A I D A Y P L S K H P N S E Y F  
Replicase 1b

AAGGTTTTTTACGTATTACTTGATTGGGTTAAGCATCTTAACAAAAATTTGAATGAGGGTGTCTTGAATCTTTTTCTGTTACACTTCTTG  
TTCAAAAAATGCATAATGAACCTAACCAATTCGTAGAATTGTTTTAACTTACTCCCACAAGAACTTAGAAAAAGACAATGTGAAGAAC  
K V F Y V L L D W V K H L N K N L N E G V L E S F S V T L L  
Replicase 1b

TAATCAAGAAGATAAGTTTTGGTGTGAAGATTTTTATGCTAGTATGTATGAAAATTCTACAATATTGCAAGCTGCTGGCTTATGTGTTGTT  
ATTAGTTCTTCTATTCAAACACACTTCTAAAAATACGATCATACATACTTTAAGATGTTATAACGTTTCGACGACCGAATACACAACAA/  
N Q E D K F W C E D F Y A S M Y E N S T I L Q A A G L C V V  
Replicase 1b

GTGGTTCACAAACTGTTCTTCGTTGTGGTGATTGTCTGCGTAAGCCTATGTTGTGCACTAAATGTGCATATGATCATGTATTTGGTACCGAC  
CACCAAGTGTGTTGACAAGAAGCAACACCACTAACAGACGCATTGCGATACAACACGTGATTTACACGTATACTAGTACATAAACCATGGCTC  
C G S Q T V L R C G D C L R K P M L C T K C A Y D H V F G T D  
Replicase 1b

CACAAGTTTATTTTGGCTATAACACCGTATGTATGTAATGCATCAGGTTGTGGTGTTAGTGATGTTAAAAAATTGTATCTTGGTGGTTTGAA  
GTGTTCAAATAAAACCGATATTGTGGCATACATACATTACGTAGTCCAACACCACAATCACTACAATTTTTTAACATAGAACCACCAAACTT  
H K F I L A I T P Y V C N A S G C G V S D V K K L Y L G G L I  
Replicase 1b

TTACTATTGTACAAATCATAAACCACAGTTGTCCTTTCCATTATGTTCTGCTGGTAATATATTTGGTTTATATAAAAAATTCAGCAACTGGTT  
AATGATAACATGTTTAGTATTTGGTGTCAACAGAAAAGGTAATACAAGACGACCATTATATAAACCAATATATTTTAAGTCGTTGACCAA  
Y Y C T N H K P Q L S F P L C S A G N I F G L Y K N S A T G  
Replicase 1b

CCTTAGATGTTGAAGTTTTTAATAGGCTTGCAACGTCTGATTGGACTGATGTTAGGGACTATAAACTTGCTAATGATGTTAAAGATACACTT  
GGAATCTACAACCTTCAAAAATTATCCGAACGTTGCAGACTAACCTGACTACAATCCCTGATATTGAACGATTACTACAATTTCTATGTGAA  
S L D V E V F N R L A T S D W T D V R D Y K L A N D V K D T L  
Replicase 1b

AGACTCTTTGCGGCTGAACTATTAAAGCTAAAGAAGAGAGTGTTAAGTCTTCTTATGCTTTTGCAACTCTTAAAGAGGTTGTTGGACCTAA  
TCTGAGAAACGCCGACTTTGATAATTTGATTTCTTCTCTCACAATTCAGAAGAATACGAAAACGTTGAGAATTTCTCCAACAACCTGGATT

R L F A A E T I K A K E E S V K S S Y A F A T L K E V V G P K  
Replicase 1b

AGAATTGCTTCTTAGTTGGGAAAGTGGTAAAGTTAAACCACCTTTGAATCGTAATTCTGTTTTACCTGTTTTCAAATAAGTAAGGACTCAA  
TCTTAACGAAGAATCAACCCTTTACCATTTCAATTTGGTGGAACTTAGCATTAAAGACAAAAGTGGACAAAAGTTTATTCATTCCTGAGTT

E L L L S W E S G K V K P P L N R N S V F T C F Q I S K D S  
Replicase 1b

AATTCCAAATAGGTGAGTTCATCTTTGAAAAGGTTGAATATGGTTCTGATACTGTTACGTATAAGTCTACTGTAACCACTAAGTTAGTTCCT  
TTAAGGTTTATCCACTCAAGTAGAACTTTTCCAACCTATACCAAGACTATGACAATGCATATTCAGATGACATTGGTGATTCAATCAAGGA

K F Q I G E F I F E K V E Y G S D T V T Y K S T V T T K L V P  
Replicase 1b

GGTATGATTTTTGTCTTAACATCTCACAATGTTCAACCTTTACGTGCACCAACTATTGCAAACCAAGAGAAGTATTCTAGCATTTATAAATT  
CCATACTAAAAACAGAATTGTAGAGTGTTACAAGTTGGAAATGCACGTGGTTGATAACGTTTGGTTCTCTTCATAAGATCGTAAATATTTAA

G M I F V L T S H N V Q P L R A P T I A N Q E K Y S S I Y K L  
Replicase 1b

GCACCCTGCTTTTAATGTCAGTGATGCATATGCTAATTTGGTTCCATATTACCAACTTATTGGTAAACAAAAGATAACTACAATACAGGGTC  
CGTGGGACGAAAATTACAGTCACTACGTATACGATTAAACCAAGGTATAATGGTTGAATAACCATTGTGTTTCTATTGATGTTATGTCCCAG

H P A F N V S D A Y A N L V P Y Y Q L I G K Q K I T T I Q G  
Replicase 1b

CTCCTGGTAGTGGAAGTCACATTGTTCCATTGGACTTGGATTGTACTATCCAGGTGCGCGTATTGTTTTTGTGCTTGCTGCCCATGCTGCT  
GAGGACCATCACCATTCAAGTGAACAAGGTAACCTGAACCTAACATGATAGTCCACGCGCATAACAAAAACAACGAACACGGGTACGACGA

P P G S G K S H C S I G L G L Y Y P G A R I V F V A C A H A A  
Replicase 1b

GTTGATTCCTTATGTGCAAAAGCTATGACTGTTTATAGCATTGATAAGTGTACTAGGATTATACCTGCAAGAGCTCGGGTTGAGTGTTATAG  
CAACTAAGGAATACACGTTTTGATACTGACAAATATCGTAACTATTCATGATCCTAATATGGACGTTCTCGAGCCCAACTCACAATATC

V D S L C A K A M T V Y S I D K C T R I I P A R A R V E C Y S  
Replicase 1b

TGGCTTTAAACCAAATAACACTAGTGCAACAATACATATTTAGCACTGTTAACGCATTACCTGAGTGAATGCTGATATTGTTGTTGTAGATG  
ACCGAAATTTGGTTTATTGTGATCACGTGTTATGTATAAATCGTGACAATTGCGTAATGGACTCACATTACGACTATAACAACAACATCTAC

G F K P N N T S A Q Y I F S T V N A L P E C N A D I V V V D  
Replicase 1b

AAGTTTCAATGTGTACAAATTATGACCTTTCTGTTATTAATCAGCGTTTATCATATAAACATATTGTTTATGTTGGTGATCCACAACAACCT  
TTCAAAGTTACACATGTTTAATACTGGAAAGACAATAATTAGTCGCAAATAGTATTTGTATAACAAATACAACCACTAGGTGTTGTTGAA

E V S M C T N Y D L S V I N Q R L S Y K H I V Y V G D P Q O L  
Replicase 1b

21/87

CCTGCACCTAGAGTAATGATTACTAAAGGTGTTATGGAGCCTGTTGATTATAACGTTGTTACTCAACGTATGTGTGCTATAGGCCCTGATC  
GGACGTGGATCTCATTACTAATGATTTCCACAATACCTCGGACAATAATATTGCAACAATGAGTTGCATACACAGATATCCGGGACTAC

P A P R V M I T K G V M E P V D Y N V V T Q R M C A I G P D  
Replicase 1b

TTTTCTTCATAAATGTTATAGATGTCCTGCTGAAATAGTTAATACAGTTTCTGAAC TTGTTTATGAGAACAAGTTTGTCCTGTTAAACCT  
AAAAGAAGTATTTACAATATCTACAGGACGACTTTATCAATTATGTCAAAGACTTGAACAAATACTCTTGTTCAAACAGGGACAATTTGGA

F L H K C Y R C P A E I V N T V S E L V Y E N K F V P V K P  
Replicase 1b

CTAGTAAACAGTGTTTTAAATCTTTTTTAAGGGTAATGTACAGTTGACAATGGCTCTAGTATTAACAGAAAGCAGCTTGAAATAGTTAA  
GATCATTTGTCACAAAATTTAGAAAAAATTCCTTACATGTCCAAC TGTACCGAGATCATAATTGTCTTTGTCGAAC TTTATCAATT

A S K Q C F K I F F K G N V Q V D N G S S I N R K Q L E I V K  
Replicase 1b

CTGTTTTAGTTAAAAATCCAAGTTGGAGTAAGGCTGTGTTTATTTCTCCTTATAATAGTCAGAATTATGTTGCTAGTAGATTTTTAGGAC  
GACAAAATCAATTTTAGGTTCAACCTCATTCCGACACAAATAAGAGGAATATTATCAGTCTTAATACAACGATCATCTAAAAATCCTG.

L F L V K N P S W S K A V F I S P Y N S Q N Y V A S R F L G  
Replicase 1b

TCAAATTCAAACTGTTGATTCTTCTCAAGGTAGTGAGTATGATTATGTAATCTATGCACAACTTCTGACACTGCACATGCTTGCAATGTA  
AGTTTAAGTTTGACAATAAGAAGAGTTCCATCACTCATACTAATACATTAGATACGTGTTTGAAGACTGTGACGTGTACGAAGGTTACAT

Q I Q T V D S S Q G S E Y D Y V I Y A Q T S D T A H A C N V  
Replicase 1b

ACCGTTTTAATGTTGCTATAACACGTGCTAAGAAGGGTATATTTTGTGTAATGTGTGATAAACTTTGTTTGATTCACTTAAGTTTTTGAC  
TGGCAAATTTACAACGATATTGTGCACGATTCTTCCCATATAAAACACATTACACACTATTTTGAACAACTAAGTGAATTCAAAAACTC

N R F N V A I T R A K K G I F C V M C D K T L F D S L K F F E  
Replicase 1b

ATTAACATGCAGATTTACTCTAGCCAGGTTTGTGGCTTGTTTAAAAATTGTACACGCACTCCTCTTAATTTACCACCAACTCATGCACA  
TAATTTGTACGTCTAAATGTGAGATCGGTCCAAACACCGAACAAATTTTAAAGATGTGCGTGAGGAGAATTAATGGTGGTTGAGTACGTGT

I K H A D L H S S Q V C G L F K N C T R T P L N L P P T H A I  
Replicase 1b

CACTTTCTTGTCGTTGTCAGATCAGTTTAAGACTACAGGTGATTTAGCTGTTCAAATAGGTTCAAATAATGT-TTGACTTATGAACATGTTA  
GTGAAAGAACAGCAACAGTCTAGTCAAATCTGATGTCCACTAAATCGACAAGTTTATCCAAGTTTATTACAAACATGAATACTTGTAAT

T F L S L S D Q F K T T G D L A V Q I G S N N V C T Y E H V  
Replicase 1b

TATCATTTATGGGTTTTAGGTTTGATATTAGTATTCCTGGTAGTCATAGTTTGTGTTTGTACACGTGACTTTGCTATTCGTAATGTGCGTGGT  
ATAGTAAATACCCAAAATCCAACTATAATCATAAGGACCATCAGTATCAAACAAAACATGTGCACTGAAACGATAAGCATTACACGCACCA

I S F M G F R F D I S I P G S H S L F C T R D F A I R N V R G  
Replicase 1b

TGGTTGGGTATGGATGTTGAAAGTGCTCATGTTTGTGGCGATAACATAGGTACTAATGTTCCCTTTACAGGTTGGTTTTTCAAATGGTGTTAA  
ACCAACCCATACCTACAACCTTTCACGAGTACAAACACCGCTATTGTATCCATGATTACAAGGAAATGTCCAACCAAAAAGTTTACCACAATT  
W L G M D V E S A H V C G D N I G T N V P L Q V G F S N G V N  
Replicase 1b

TTTTGTGTGCAAACTGAAGGTTGTGTGTCTACCAATTTTGGTGATGTTATTAACCTGTTTGTGCAAAATCTCCACCAGGTGAACAATTTA  
AAAAACAACAGTTTGACTTCCAACACACAGATGGTTAAAACCACTACAATAATTTGGACAAACACGTTTTAGAGGTGGTCCACTTGTTAAAT  
F V V Q T E G C V S T N F G D V I K P V C A K S P P G E Q F  
Replicase 1b

GACACCTTGTTCCCTTTTTACGTAAAGGACAACCTTGGTTAATTGTTTCGTAGACGCATTGTGCAAAATGATATCTGATTATTTGTCCAATTTG  
CTGTGGAACAAGGAAAAAATGCATTTCCTGTTGGAACCAATTAACAAGCATCTGCGTAACACGTTTACTATAGACTAATAAACAGGTAAAC  
R H L V P F L R K G Q P W L I V R R R I V Q M I S D Y L S N L  
Replicase 1b

TCTGACATTCTTGTCTTTGTTTTGTGGGCAGGTAGTTTGGAAATTAACAATGCGTTACTTTGTAAAAATAGGGCCAATTAATATTGTTA  
AGACTGTAAGAACAGAAACAAAACACCGTCCATCAAACCTTAATTGATGTTACGCAATGAAACATTTTTATCCCGGTTAATTTATAACAAT  
S D I L V F V L W A G S L E L T T M R Y F V K I G P I K Y C Y  
Replicase 1b

TTGTGGTAATTCTGCCACTTGTTATAATTCAGTTAGTAATGAATATTGTTGTTTTAAACATGCATTGGGTTGTGATTATGTTTACAATCCGT  
AACACCATTAAGACGGTGAACAATATTAAGTCAATCATTACTTATAACAACAAAATTTGTACGTAACCCAACACTAATACAAATGTTAGGCA  
C G N S A T C Y N S V S N E Y C C F K H A L G C D Y V Y N P  
Replicase 1b

ATGCTTTTGATATACAACAGTGGGGTTATGTTGGTTCCTTGAGCCAGAACCACCACACGTTCTGTAACATTTCATAGAAACGAGCATGATGCT  
TACGAAAACATATGTTGTCACCCCAATACAACCAAGGAACCTCGGTCTTGGTGGTGTGCAAGACATTGTAAGTATCTTTGCTCGTACTACGA  
Y A F D I Q Q W G Y V G S L S Q N H H T F C N I H R N E H D A  
Replicase 1b

TCTGGTGATGCTGTTATGACACGTTGTTTGGCAGTACATGATTGTTTTGTCAAAAATGTTGATTGGACTGTAACGTACCCCTTTATTGCAAA  
AGACCACTACGACAATACTGTGCAACAAACCGTCATGTACTAACAAACAGTTTTTACAACCTAACCTGACATTGCATGGGGAAATAACGTTT  
S G D A V M T R C L A V H D C F V K N V D W T V T Y P F I A N  
Replicase 1b

TGAGAAATTTATCAATGGCTGTGGGCGTAATGTCCAGGGACATGTTGTTTCGCGCAGCCTTGAAATTGTATAAACCTAGTGTTATTCATGATA  
ACTCTTTAAATAGTTACCGACACCCGATTACAGGTCCCTGTACAACAGCGCTCGGAACCTTAACATATTTGGATCACAATAAGTACTAT  
E K F I N G C G R N V Q G H V V R A A L K L Y K P S V I H D  
Replicase 1b

TTGGTAATCCTAAAGGTGTACGTTGTGCTGTTACTGATGCCAAATGGTACTGTTATGACAAGCAACCTGTTAATAGTAATGTCAAGTTGTTG  
AACCATTAGGATTTCCACATGCAACACGACAATGACTACGGTTTACCATGACAATACTGTTTCGTTGGACAATTATCATTACAGTTCAACAAC  
I G N P K G V R C A V T D A K W Y C Y D K Q P V N S N V K L L  
Replicase 1b

GATTATGATTATGCAACCCATGGTCAACTTGATGGTCTTTGTTTATTCTGGAATTGTAATGTTGATATGTATCCAGAATTTTCAATTGTGT  
CTAATACTAATACGTTGGGTACCAAGTTGAACTACCAGAAACAATAAGACCTTAACATTACAACATACATAGGTCTTAAAAGTTAACACA  
D Y D Y A T H G Q L D G L C L F W N C N V D M Y P E F S I V  
Replicase 1b

TCGCTTTGACACACGTAAGTCTGTTCTGTTTTAATTTAGAAGGTGTTAATGGTGGTCTCTTTATGTTAACAACATGCGTTTCATACACCA  
AGCGAAACTGTGTGCATGAGCAAGACAAAAATTAATCTTCCACAATTACCACCAAGAGAAATACAATTGTTTGTACGCAAAGTATGTGGT  
R F D T R T R S V F N L E G V N G G S L Y V N K H A F H T P  
Replicase 1b

CATATGATAAACGTGCTTTTGTAAATTAACCTATGCCCTTTTTTACTTTGATGACAGTGATTGTGATGTTGTGCAAGAACAAGTTAA  
GTATACTATTTGCACGAAAACAATTTAATTTTGGATACGGGAAAAAAATGAACTACTGTCTACTAACACTACAACAGTTCTTGTTCATT.  
A Y D K R A F V K L K P M P F F Y F D D S D C D V V Q E Q V N  
Replicase 1b

TATGTACCCCTTCGCGCTAGTAGTTGTGTTACCCGTTGTAATATAGGTGGTGTCTTTGTTCAAACATGCAAATTTGTATCAAAAATATG  
ATACATGGGGAAGCGCATCATCAACACAATGGGCAACATTATATCCACCACGACAAACAAGTTTTGTACGTTTAAACATAGTTTTTATAC,  
Y V P L R A S S C V T R C N I G G A V C S K H A N L Y Q K Y  
Replicase 1b

TGAGGCATATAATACATTTACACAGGCTGGTTTTAACATTTGGGTACCACATAGTTTTGATGTTTATAATTTGTGGCAAATTTTTATTGAA/  
ACTCCGTATATTATGTAATGTGTCCGACCAAAATTGTAAACCCATGGTGTATCAAACTACAAATATTAACACCGTTTAAAAATAACTT.  
E A Y N T F T Q A G F N I W V P H S F D V Y N L W Q I F I E  
Replicase 1b

CTAATTTACAAAGTCTTGAAAATATAGCATTAAATGTTGTAAGGAGGTTTACTGGTGTGATGGTGAGTTACCTGTTGCAGTTGT  
GATTAAATGTTTCAGAACTTTTATATCGTAAATTACAACATTTTTTCCACAAAATGACCACAACCTACCACTCAATGGACAACGTCAACA  
T N L Q S L E N I A F N V V K K G C F T G V D G E L P V A V V  
Replicase 1b

AACGACAAAGTTTTTGTTCGCTATGGCGATGTTGACAACTTGGTTTTTACAAATAAAACAACATTGCCTACTAATGTTGCTTTTGAATTGTT  
TTGCTGTTTCAAAAACAAGCGATACCGCTACAACCTGTTGAACCAAAAATGTTTATTTTGTGTAACGGATGATTACAACGAAAACCTTAACAA  
N D K V F V R Y G D V D N L V F T N K T T L P T N V A F E L F  
Replicase 1b

TGCAAAACGAAAAATGGGTTTAAACACCACCATTGTCTATTTCICAAAAATCTGGTGTGTTGCTACATATAAATTTGTTTATGGGATTATG  
ACGTTTTGCTTTTACCAAATTTGGGTGGTAACAGATAAGAGTTTTTAGAACCAACAACGATGTATATTTAAACAAAATACCCTAATAC  
A K R K M G L T P P L S I L K N L G V V A T Y K F V L W D Y  
Replicase 1b

AAGCTGAAAGACCTTTTACCTCATATACTAAGAGTGTATGTAAATACACTGATTTTAATGAGGATGTTTGTGTTTGTGTTTGAATAGTATT  
TTCGACTTTCTGAAAAATGGAGTATATGATTCTCACATACATTTATGTGACTAAAATTACTCCTACAAACACAAACAAAACCTGTTATCATAA  
E A E R P F T S Y T K S V C K Y T D F N E D V C V C F D N S I  
Replicase 1b

CAGGGTTCGTATGAGCGTTTTACGCTTACTACGAACGCTGTTTTATTTTCTACTGTTGTCATTAAAAATTTAACACCTATAAAGTTGAATTT  
GTCCCAAGCATACTCGCAAAATGCGAATGATGCTTGCACAAAATAAAAGATGACAACAGTAATTTTAAATTGTGGATATTTCAACTTAAA

Q G S Y E R F T L T T N A V L F S T V V I K N L T P I K L N F  
Replicase 1b

TGGTATGTTGAATGGTATGCCAGTTTCTTCTATTAAGAGTGATAAAGGTGTTGAAAAATTAGTTAATTGGTACACATATGTTTCGTAAAAATG  
ACCATACAACCTACCATACGGTCAAAGAAGATAATTCTCACTATTTCCACAACCTTTTAATCAATTAACCATGTGTATACAAGCATTTTAC

G M L N G M P V S S I K S D K G V E K L V N W Y T Y V R K N  
Replicase 1b

GTCAATTTCAAGATCATTATGATGGTTTTTACACTCAAGGTAGGAATTTATCAGACTTTACACCAAGAAGTGATATGGAGTATGATTTTCTT  
CAGTTAAAGTTCTAGTAATACTACCAAAAATGTGAGTTCCATCCTTAAATAGTCTGAAATGTGGTTCTTCACTATACCTCATACTAAAAGAA

G Q F Q D H Y D G F Y T Q G R N L S D F T P R S D M E Y D F L  
Replicase 1b

AACATGGATATGGGTGTTTTTATTAATAAATATGGTCTTGAGGATTTTAATTTGAACATGTTGTATATGGTGATGTTTCAAAAACCTACATT  
TTGTACCTATACCCACAAAAATAATTATTTATACCAGAACTCCTAAAATTAACCTGTACAACATATACCACTACAAAGTTTTTGATGTAA

N M D M G V F I N K Y G L E D F N F E H V V Y G D V S K T T L  
Replicase 1b

AGGAGGTCTTCATTTGTTGATATCACAGTTTAGGCTTAGTAAAATGGGTGTTTTGAAAGCTGATGATTTTGTCACTGCTTCTGACACAACCTT  
TCCTCCAGAAGTAAACAACCTATAGTGCAATCCGAATCATTTTACCCACAAAACCTTCGACTACTAAACAGTGACGAAGACTGTGTTGAA

G G L H L L I S Q F R L S K M G V L K A D D F V T A S D T T  
Replicase 1b

TGAGGTGCTGTACTGTTACTTATCTTAATGAACCTTAGTTCAAAAGTTGTTTGTACTTATATGGATTTGTTGTTGGACGACTTTGTTACTATA  
ACTCCACGACATGACAATGAATAGAATTACTTGAATCAAGTTTCAACAAACATGAATATACCTAAACAACAACCTGCTGAAACAATGATAT

L R C C T V T Y L N E L S S K V V C T Y M D L L L D D F V T I  
Replicase 1b

CTAAAGAGTTTAGATCTTGGTGTAATATCTAAAGTTCATGAAGTTATTATAGATAATAAACCTTATAGGTGGATGTTGTGGTGTAAGATAA  
GATTTCTCAAATCTAGAACCACATTATAGATTTCAAGTACTTCAATAATATCTATTATTTGGAATATCCACCTACAACACCACATTTCTATT

L K S L D L G V I S K V H E V I I D N K P Y R W M L W C K D N  
Replicase 1b

CCACTTGTCGACTTTTTATCCACAGTTGCAGTCTGCTGAATGGAAGTGTTGTTATGCTATGCCACAAATTTATAAGCTTCAACGTATGTGTT  
GGTGAACAGCTGAAAAATAGGTGTCAACGTCAGACGACTTACCTTCACACCAATACGATACGGTGTTTAAATATTCGAAGTTGCATACACAA

H L S T F Y P Q L Q S A E W K C G Y A M P Q I Y K L Q R M C  
Replicase 1b

TGGAACCTTGTAATTTATATAATTATGGTGCTGGTATTAAGTTGCCTAGTGGTATAATGTTAAATGTTGTTAAATACACTCAGCTTTGTCAA  
ACCTTGGAACATTAAATATATTAATACCACGACCATAATTCAACGGATCACCATATTACAATTTACAACAATTTATGTGAGTCGAAACAGTT

L E P C N L Y N Y G A G I K L P S G I M L N V V K Y T Q L C Q  
Replicase 1b



TACCTAAATAGCACTACAATGTGCGTACCTCATAATATGCGTGTGTTTGCCTATGGTGCTGGTTCTGACAAAGGTGTGGCACCTGGTACAA  
ATGGATTTATCGTGATGTTACACGCATGGAGTATTATACGCACAAAACGTGATACCACGACCAAGACTGTTTCCACACCGTGGACCATGTT  
Y L N S T T M C V P H N M R V L H Y G A G S D K G V A P G T  
Replicase 1b

TGTTTTAAACGTTGGCTACCACCTGATGCAATAATCATTGATAATGATATCAATGATTATGTTAGTGATGCAGATTTTAGCATTACAGGT  
ACAAAATTTTGAACCGATGGTGGACTACGTTATTAGTAACTATTACTATAGTTACTAATAACAATCACTACGTCTAAAATCGTAATGTCCA  
V L K R W L P P D A I I I D N D I N D Y V S D A D F S I T G  
Replicase 1b

ATTGTGCTACTGTTTACCTTGAAGATAAGTTTGACTTACTTATTTCTGATATGTATGATGGTAGAATTAATTTTGTGATGGTGAAAACGT  
TAACACGATGACAAATGGAACCTTCTATTCAAACCTGAATGAATAAAGACTATACATACTACCATCTTAATTTAAACACTACCACTTTTGCA  
D C A T V Y L E D K F D L L I S D M Y D G R I K F C D G E N V  
Replicase 1b

TCTAAAGATGGTTTTTTTACTTATCTTAATGGTGTATTAGAGAAAAATTAGCTATTGGTGGTAGTGTGGCCATTAAGATTACAGAATATAG  
AGATTTCTACCAAAAAAATGAATAGAATTACCACAATAATCTCTTTTAAATCGATAACCACCATCACAACGGTAATTCTAATGTCTTATATC  
S K D G F F T Y L N G V I R E K L A I G G S V A I K I T E Y :  
Replicase 1b

TTGGAATAAGTATCTTTATGAATTAATACAAAGATTTGCTTTTTGGACTTTGTTCTGCACGTCTGTTAATACATCCTCTTCAGAAGCTTTTC  
AACCTTATTCATAGAAATACTTAATTATGTTTCTAAACGAAAAACCTGAAACAAGACGTGCAGACAATTATGTAGGAGAAGTCTTCGAAAAG  
W N K Y L Y E L I Q R F A F W T L F C T S V N T S S S E A F  
Replicase 1b

TTATTGGTATTAATTATTTAGGTGACTTTATTCAAGGTCCTTTTATAGCTGGTAACACTGTTTCATGCTAATTATATATTTTGGCGTAATTCT  
ATAACCATTAATTAATAATCCACTGAAATAAGTTCCAGGAAAATATCGACCATTGTGACAAGTACGATTAATATATAAAACCGCATTAAGA  
L I G I N Y L G D F I Q G P F I A G N T V H A N Y I F W R N S  
Replicase 1b

ACTATTATGCTTTTGTACATAAATTCAGTTTTAGATTTAAGTAAGTTGAATGTAAACATAAGGCCACTGTTGTTGTTACACTTAAAGATAG  
TGATAATACAGAAACAGTATGTTAAGTCAAAATCTAAATTCATTCAAACCTTACATTTGTATTCCGGTGACAACAACAATGTGAATTTCTATC  
T I M S L S Y N S V L D L S K F E C K H K A T V V V T L K D S  
Replicase 1b

TGATGTAAATGATATGGTTTTGAGTTTGATTAAGAGTGGTAGGTTGTTGTTACGTAATAGTGGCCGTTTGGTGGTTTGTAGTAATCATTAG  
ACTACATTTACTATACCAAAACTCAAACCTAATCTCACCATCCAACAACAATGCATTATCACCGGCAAAACCACCAAAATCATTAGTAAATC  
D V N D M V L S L I K S G R L L L R N S G R F G G F S N H L  
Replicase 1b

TCTCAACTAAATGAAACTTTTCTTGATTTTGCTTATTTTGCCCTGGTTTCTTGCTTTTCTACATGTAACAGTAATGCTAGTATTTCTATGT  
AGAGTTGATTACTTTGAAAAGAACTAAAACGAATAAAACGGGGACCAAGAACGAAAAGATGTACATTGTCATTACGATCATAAAGATACA

M K L F L I L L I L P L V S C F S T C N S N A S I S M  
Spike

V S T K  
— Replicase 1b —

TACAATTAGGTGTTCTGATAACTCTTCAACTATTGTCACAGTTTGTGTCAGTCCATTGGATTGTGCTAATCAGAGTACATCTAGTTAC  
ATGTTAATCCACAAGGACTATTGAGAAGTTGATAACAGTGTCCAAACAACGGTCAGGTAACCTAAACACGATTAGTCTCATGTAGATCAATG

L Q L G V P D N S S T I V T G L L P V H W I C A N Q S T S S Y  
Spike

CCAGCCAACGGCTTTTCTATATTGATGTTGGTAAACACCGTAGTGCCTTGCACCTCCATAGTGGTTATTATGATGCTAACCAGTATTATAT  
GGTCGGTTGCCGAAAAGATATAACTACAACCATTTGTGGCATCACGGAACGTGAGGTATACCAATAATACTACGATTGGTCATAATATA

P A N G F F Y I D V G K H R S A F A L H S G Y Y D A N Q Y Y I  
Spike

TTATCTCACTAATAAAATACATTTAAATGCTCCTGTCACTCTGAAGATTTGTAAGTTGGAAACACTTCTTTTGATTTTTTAAGTAATGTTT  
AATAGAGTGATTATTTATGTAAATTTACGAGGACAGTGAGACTTCTAAACATTCAAACCTTTGTGAAGAAAACATAAAATTCATTACAAA

Y L T N K I H L N A P V T L K I C K F G N T S F D F L S N V  
Spike

CTACTTCTCATGATTGTATAGTTAATTTGTCATTCACAGAACAGTTAGGTGTGCCTTTGGGCATAACTATATCGGGTGAAACTGTACGTTTG  
GATGAAGAGTACTAACATATCAATTAACAGTAAGTGTCTTGTCAATCCACACGGAACCCGTATTGATATAGCCACTTTGACATGCAAAC

S T S H D C I V N L S F T E Q L G V P L G I T I S G E T V R L  
Spike

CATTTATATAATGCAACTCGTACTTTTATGTGCCGGCCGCTTATAAACTTACTAAACTTAGTGTTAAATGTTACTTTAGTGAATCCTGTGT  
GTAAATATATTACGTTGAGCATGAAAAATACACGGCCGGCGAATATTTGAATGATTGAATCACAATTTACAATGAAATCACTTAGGACACA

H L Y N A T R T F Y V P A A Y K L T K L S V K C Y F S E S C V  
Spike

TTTTAGTGTGTCAATGCCACCATTACTGTTAATGTCACCACACTTAATGGCCGTATAGTTAACTACACTGTTTGTGATGATTGTAATGGTT  
AAAATCACAACAGTTACGGTGGAATGACAATTACAGTGGTGTGAATTACCGGCATATCAATTGATGTGACAAACACTACTAACATTACCAA

F S V V N A T I T V N V T T L N G R I V N Y T V C D D C N G  
Spike

ATACTGATAACATATTTTCTGTTCAACAGGATGGCCGCATTCTAATGGTTTCCCTTTTAATAATTGGTTTTGTAACTAATGGTTCCACA  
TATGACTATTGTATAAAGACAAGTTGTCCTACCGGCGTAAGGATTACCAAAGGGAAAATTATTAACCAAAAACAATTGATTACCAAGGTGT

Y T D N I F S V Q Q D G R I P N G F P F N N W F L L T N G S T  
Spike

## EMCR-CoV.MPD (1 &gt; 27532) Site and Sequence

TTAGTGGACGGGGTCTCTAGACTTTATCAACCACTCCGTTTAACTTGTGTTATGGCCTGTACCTGGTCTTAAATCTTCAACTGGTTTTGTT  
AATCACCTGCCCCAGAGATCTGAAATAGTTGGTGAGGCAAATTGAACAAATACCGGACATGGACCAGAATTTAGAAGTTGACCAAAACAA/  
L V D G V S R L Y Q P L R L T C L W P V P G L K S S T G F V  
Spike

TTTTAATGCCACTGGTTCTGATGTTAATTGTAACGGCTATCAACATAATTCTGTTGCTGATGTTATGCGTTACAATCTTAACCTCAGTGCT  
AAAATTACGGTGACCAAGACTACAATTAACATTGCCGATAGTTGTATTAAAGACAACGACTACAATACGCAATGTTAGAATTGGAGTCACGA  
F N A T G S D V N C N G Y Q H N S V A D V M R Y N L N L S A  
Spike

ATTCTGTGGACAATCTTAAGAGTGGTGTATAGTTTTTAAACTTTACAGTACGATGTTTTGTTTTATTGTAGTAATCTTCTTCAGGTGT  
TAAGACACCTGTTAGAATTCTCACCACAATATCAAAAATTTTGAAATGTCATGCTACAAAACAAAATAACATCATTAAGAAGAAGTCCACA  
N S V D N L K S G V I V F K T L Q Y D V L F Y C S N S S S G V  
Spike

CTTGACACCACAATACCTTTTGGCCCTTCCTCTCAACCTTATTACTGTTTTATAAACAGTACTATCAACACTACTCATGTTAGCACTTTTG  
GAACTGTGGTGTATGGAAAACCGGAAGGAGAGTTGGAATAATGACAAAATATTTGTCATGATAGTTGTGATGAGTACAATCGTGAAAAC  
L D T T I P F G P S S Q P Y Y C F I N S T I N T T H V S T F  
Spike

GGGTATTTTACCACCCACTGTGCGTGAAATTGTTGTTGCTAGAACTGGTCAGTTTTATATTAATGGTTTTAAGTATTTGATTTGGGTTTC/  
CCATAAAATGGTGGGTGACACGCACCTTAAACAACACGATCTTGACCAGTCAAAATATAATTACCAAATTCATAAAGCTAAACCCAAAG  
G I L P P T V R E I V V A R T G Q F Y I N G F K Y F D L G F  
Spike

TAGAAGCTGTCAATTTTAATGTCACGACTGCTAGTGCCACAGATTTTTGGACGGTTGCATTTGCTACTTTTGTGATGTTTTGGTTAATGTI  
ATCTTCGACAGTTAAAATTACAGTGCTGACGATCACGGTGTCTAAAAACCTGCCAACGTAAACGATGAAAACAACCTACAAAACCAATTACAA  
I E A V N F N V T T A S A T D F W T V A F A T F V D V L V N V  
Spike

AGTGCAACTAACATTCAAACTTACTTTATTGCGATTCTCCATTTGAAAAGTTGCAGTGTGAGCACTTGCAGTTTGGATTGCAAGATGGTTT  
TCACGTTGATTGTAAGTTTGAATGAAATAACGCTAAGAGGTAAACTTTTCAACGTCACACTCGTGAACGTCAAACCTAACGTTCTACCAAA  
S A T N I Q N L L Y C D S P F E K L Q C E H L Q F G L Q D G F  
Spike

TTATTCTGCAAATTTCTTGATGATAATGTTTTGCCTGAGACTTATGTTGCACTCCCATTTATTATCAACATACGGACATAAATTTTACTG  
AATAAGACGTTTAAAGAAGTACTATTACAAAACGGACTCTGAATACAACGTGAGGGGTAAATAATAGTTGTATGCCTGTATTTAAATGAC  
Y S A N F L D D N V L P E T Y V A L P I Y Y Q H T D I N F T  
Spike

CAACTGCATCTTTTGGTGGTTCTTGTTATGTTTGTAACCACGCCAGGTTAATATATCTCTTAATGGTAACACTTCAGTGTGTGTTAGAACA  
GTTGACGTAGAAAACCAAGAACAATACAAACATTTGGTGCGGTCCAATTATATAGAGAATTACCATTGTGAAGTCACACACAATCTTGT  
A T A S F G G S C Y V C K P R Q V N I S L N G N T S V C V R T  
Spike

TCTCATTTTTCAATTAGGTATATTTATAACCGCGTTAAGAGTGGTTCACCAGGTGACTCTTCATGGCATATTTATTTAAAGAGTGGCACTTG  
AGAGTAAAAAGTTAATCCATATAAATATTGGCGCAATTCTCACCAAGTGGTCCACTGAGAAGTACCGTATAAATAAATTTCTCACCCTGAAC

S H F S I R Y I Y N R V K S G S P G D S S W H I Y L K S G T C  
Spike

TCCATTTTCTTTTCTAAGTTAAATAATTTTCAAAAGTTTAAGACTATTTGTTTCTCAACCGTCGAAGTGCCTGGTAGTTGTAATTTTCCAC  
AGGTAAAAGAAAAAGATTCAATTTATTAAGTTTTCAAATTCTGATAAACAAAGAGTTGGCAGCTTCACGGACCATCAACATTAAGAGTG

P F S F S K L N N F Q K F K T I C F S T V E V P G S C N F P  
Spike

TTGAAGCCACCTGGCATTACACTTCTTATACTATTGTTGGTGCTTTGTATGTTACTTGGTCTGAAGGTAATTCATTACTGGTGTACCTTAT  
AACTTCGGTGGACCGTAATGTGAAGAATATGATAACAACCACGAAACATACAATGAACCAGACTTCCATTAAGGTAATGACCACATGGAATA

L E A T W H Y T S Y T I V G A L Y V T W S E G N S I T G V P Y  
Spike

CCTGTCTCTGGTATTCGTGAGTTTAGTAATTTAGTTTTAAATAATTGTACCAAATATAATATTTATGATTATGTTGGTACTGGAATTATACG  
GGACAGAGACCATAAGCACTCAAATCATTAAATCAAATTTATTAACATGGTTTATATTATAAATACTAATAACAACCATGACCTTAATATGC

P V S G I R E F S N L V L N N C T K Y N I Y D Y V G T G I I R  
Spike

TTCTTCAAACCAAGTCACTTGCTGGTGGTATTACATATGTTTCTAACTCTGGTAATTTACTTGGTTTTAAAAATGTTTCCACTGGTAACATTT  
AAGAAGTTTGGTCAGTGAACGACCACCATAATGTATACAAAGATTGAGACCATTAAATGAACCAAAATTTTACAAAGGTGACCATTGTAAA

S S N Q S L A G G I T Y V S N S G N L L G F K N V S T G N I  
Spike

TTATTGTGACACCATGTAACCAACCAGATCAAGTAGCTGTTTATCAACAAAGCATTATTGGTGCCATGACCGCTGTTAATGAGTCTAGATAT  
AATAACACTGTGGTACATTGGTTGGTCTAGTTCATCGACAAATAGTTGTTTCGTAATAACCACGGTACTGGCGACAATTACTCAGATCTATA

F I V T P C N Q P D Q V A V Y Q Q S I I G A M T A V N E S R Y  
Spike

GGCTTGCAAACTTACTACAGTTACCTAACTTTTATTATGTTAGTAATGGTGGTAACAATTGCACTACGGCTGTTATGATTTATTCTAATTT  
CCGAACGTTTTGAATGATGTCAATGGATTGAAAATAATACAATCATTACCACCATTGTAAACGTGATGCCGACAATACTAAATAAGATTAAA

G L Q N L L Q L P N F Y Y V S N G G N N C T T A V M I Y S N F  
Spike

TGGTATTTGTGCTGATGGTTCTTTAATTCCTGTTCTGTCGCGTAATTCAGTGATAATGGTATTTTACGCCATAATCACTGCTAATTTATCCA  
ACCATAAACACGACTACCAAGAAATTAAGGACAAGCAGGCGCATTAAAGTCACTATTACCATAAAGTCGGTATTAGTGACGATTAAATAGGT

G I C A D G S L I P V R P R N S S D N G I S A I I T A N L S  
Spike

TTCCCTCTAACTGGACTACTTCAGTTCAAGTTGAGTACCTCCAAATTACTAGTACTCCAATAGTTGTTGATTGTGCTACTTATGTGTGTAAT  
AAGGGAGATTGACCTGATGAAGTCAAGTTCAACTCATGGAGGTTTAAATGATCATGAGGTTATCAACAACAAACACGATGAATACACACATTA

I P S N W T T S V Q V E Y L Q I T S T P I V V D C A T Y V C N  
Spike

GGTAACCCTCGTTGTAAGAATCTACTTAAGCAGTATACTTCTGCTTGTAAAACATTGAAGATGCCTTACGACTTAGTGCTCATTGGA  
CCATTGGGAGCAACATTCTTAGATGAATTCGTCATATGAAGACGAACATTTTGATAACTTCTACGGAATGCTGAATCACGAGTAAACCTTT

G N P R C K N L L K Q Y T S A C K T I E D A L R L S A H L E  
Spike

TAATGATGTTAGTAGTATGCTAACTTTGATAGCAATGCTTTTAGTTTGGCTAATGTTACTAGTTTTGGAGATTATAACCTTTCTAGTGTT  
ATTACTACAATCATCATACGATTGAAAGCTATCGTTACGAAATCAAACCGATTACAATGATCAAAACCTCTAATATTGGAAAGATCACAA

N D V S S M L T F D S N A F S L A N V T S F G D Y N L S S V  
Spike

TACCTCAGAGAAACATTCATTCAAGCCGTATAGCAGGACGTAGTGCTTTGGAAGATTTGTTGTTTAGCAAAGTTGTTACATCTGGTTTGGG  
ATGGAGTCTCTTTGTAAGTAAGTTCGGCATATCGTCCTGCATCACGAAACCTTCTAAACAACAAATCGTTTCAACAATGTAGACCAAACCC

L P Q R N I H S S R I A G R S A L E D L L F S K V V T S G L G  
Spike

ACTGTTGATGTTGACTATAAGTCTTGTACTAAAGGTCTTTCTATTGCTGACCTTGCTTGTGCTCAGTACTACAATGGCATAATGGTTTGGC  
TGACAACTACAAGTATTCAGAACATGATTTCCAGAAAGATAACGACTGGAACGAACACGAGTCATGATGTTACCGTATTACCAAACG

T V D V D Y K S C T K G L S I A D L A C A Q Y Y N G I M V L  
Spike

AGGTGTTGCTGATGCTGAACGTATGGCCATGTACACAGGTTCTCTTATAGGTGGCATGGTGCTCGGAGGTCTTACATCAGCAGCCGCATAC  
TCCACAACGACTACGACTTGATACCGGTACATGTGTCCAAGAGAATATCCACCGTACCACGAGCCTCCAGAATGTAGTCGTCGGCGGTATC

G V A D A E R M A M Y T G S L I G G M V L G G L T S A A A I  
Spike

CTTTTTCTTTGGCACTGCAAGCAGCACTTAAGTATGTTGCTTTACAACTGATGTGCTTCAAGAAAATCAGAAAATTTTGGCTGCATCATT  
GAAAAAGAAACCGTGACGTTCTGTGCTGAATTGATACAACGAAATGTTTGACTACACGAAGTTCTTTTAGTCTTTTAAACCGACGTAGTAAA

P F S L A L Q A R L N Y V A L Q T D V L Q E N Q K I L A A S F  
Spike

AATAAGGCTATTAATAATATTGTTGCTTCTTTTAGTAGCGTTAATGATGCTATTACACATACTGCAGAGGCTATACATACTGTTACTATTGC  
TTATTCCGATAATTATTATAACAACGAAGAAAATCATCGCAATTACTACGATAATGTGTATGACGTCTCCGATATGTATGACAATGATAACG

N K A I N N I V A S F S S V N D A I T H T A E A I H T V T I A  
Spike

ACTTAATAAGATTCAGGATGTTGTTAATCAACAGGGTAGTGCTCTTAACCATCTCACTTCACAATTGAGACATAATTTTCAAGGCCATTTCTA  
TGAATTATTCTAAGTCTTACAACAATTAGTTGTCCCATCACGAGAATTGGTAGAGTGAAGTGTTAACTCTGTATTAAAAGTCCGGTAAAGAT

L N K I Q D V V N Q Q G S A L N H L T S Q L R H N F Q A I S  
Spike

ATTCAATTCATGCTATTTATGACCGGCTTGATTCAATTCAAGCCGATCAACAAGTTGACAGATTAATTACTGGACGGCTTGACGCTTTGAAT  
TAAGTTAAGTACGATAAATACTGGCCGAACATAAGTTAAGTTCGGCTAGTTGTTCAACTGTCTAATTAATGACCTGCCGAACGTGAAACCTTA

N S I H A I Y D R L D S I Q A D Q Q V D R L I T G R L A A L N  
Spike

GCATTTGTTTCCCAAGTTTTGAATAAATATACTGAAGTTCGTGGTTCCAGACGCTTAGCACAGCAGAAGATTAATGAATGTGTCAAGTCACA  
CGTAAACAAAGGGTTCAAAACTTATTTATATGACTTCAAGCACCAAGGTCTGCGAATCGTGTCTTCTAATTACTTACACAGTTCAGTGT

A F V S Q V L N K Y T E V R G S R R L A Q Q K I N E C V K S O  
Spike

ATCTAATAGATATGGTTTTGTGGCAATGGCACTCACATCTTTTCAATCGTCAACTCAGCTCCAGATGGTTTGCTTTTTCTTCATACTGTTT  
TAGATTATCTATACCAAAACACCGTTACCGTGAGTGTAGAAAAGTTAGCAGTTGAGTCGAGGTCTACCAACGAAAAAGAAGTATGACAAA

S N R Y G F C G N G T H I F S I V N S A P D G L L F L H T V  
Spike

TGCTGCCAACTGATTACAAGAATGTAAAGGCGTGGTCTGGTATCTGTGTTGATGGCATTATGGCTATGTTCTGCGTCAACCTAACTTGGTT  
ACGACGGTTGACTAATGTTCTTACATTTCCGCACCAGACCATAGACACAACCTACCGTAAATACCGATACAAGACGCAGTTGGATTGAACCAA

L L P T D Y K N V K A W S G I C V D G I Y G Y V L R Q P N L V  
Spike

CTTTATTCTGATAATGGTGTCTTTCTGTGTAACCTCCAGGGTCATGTTTCAACCTCGTTTACCTGTTTTGTCTGATTTTGTGCAAAATATATAA  
GAAATAAGACTATTACCACAGAAAGCACATTGAAGGTCCCAGTACAAAGTTGGAGCAAATGGACAAAACAGACTAAAACAGTTTATATATT

L Y S D N G V F R V T S R V M F Q P R L P V L S D F V Q I Y N  
Spike

TTGTAATGTTACTTTTGTAAACATATCTCGTGTGAGTTACATACTGTCATACCTGACTACGTTGATGTTAATAAAACATTACAAGAGTTTG  
AACATTACAATGAAAACAATTGTATAGAGCACAGCTCAATGTATGACAGTATGGACTGATGCAACTACAATTATTTGTAATGTTCTCAAAC

C N V T F V N I S R V E L H T V I P D Y - V D V N K T L Q E F  
Spike

CACAAAACCTACCAAAGTATGTTAAGCCTAATTTTGACTTGACTCCTTTTAATTTAACATATCTTAATTTGAGTTCTGAGTTGAAGCAACTC  
GTGTTTTGAATGGTTTCATACAATTCGGATTAAAACTGAAGTGAAGGAAATTAATTTGTATAGAATTAAGTCAAGACTCAACTTCGTTGAG

A O N L P K Y V K P N F D L T P F N L T Y L N L S S E L K Q L  
Spike

GAAGCTAAAACCTGCTAGTCTTTTCCAACTACTGTTGAATTACAAGGTCTTATTGATCAGATTAACAGTACATATGTTGATTTGAAGTTGCT  
CTTCGATTTTGACGATCAGAAAAGGTTTGATGACAACCTAATGTTCCAGAATAACTAGTCTAATTGTCATGTATACAACCTAACTTCAACGA

E A K T A S L F Q T T V E L Q G L I D Q I N S T Y V D L K L L  
Spike

TAATAGGTTTGAAAATTATATCAAAATGGCCTTGGTGGGTTTGGCTCATTATTTCTGTTGTTTTGTTGATTGTTGAGTCTTCTTGTTT  
ATTATCCAACTTTTAATATAGTTTACCGGAACCAACCAACCGAGTAATAAAGACAACAAAAACATAACAACCTCAGAAGAACACAAAA

N R F E N Y I K W P W W V W L I I S V V F V V L L S L L V F  
Spike

GTTGTCTTCTACAGGTTGTTGTGGTTGTTGCAATTGTTTAACTTCATCAATGCGAGGCTGTTGTGATTGTGGTTCAACTAACTTCCTTAT  
CAACAGAAAGATGTCCAACAACCAACAACGTTAACAATGAAGTAGTTACGCTCCGACAACACTAACACCAAGTTGATTGGAAGGAATA

C C L S T G C C G C C N C L T S S M R G C C D C G S T K L P Y  
Spike

TATGAATTTGAAAAGGTCCACGTTCAATAATGCCTTTCCGGTGGCCTATTTCAACTTACTCTTGAAAGTACTATTAATAAGAGTGTGGCTAA  
ATACTTAACTTTTCCAGGTGCAAGTTATTACGGAAAGCCACCGGATAAAGTTGAATGAGAACTTTTCATGATAATTATTCTCACACCGATT

Y E F E K V H V Q .  
Spike

M P F G G L F Q L T L E S T I N K S V A N  
ORF 4ab

CTCAAATTACCACCTCATGATGTTACTGTCTTGCGTGACAATCTTAAACCTGTTACTACACTTAGTACTATCACTGCTTATTTGTTAGTTAI  
GAGTTTAATGGTGGAGTACTACAATGACAGAACGCCTGTTAGAATTTGGACAATGATGTGAATCATGATAGTGACGAATAACAATCAATI

L K L P P H D V T V L R D N L K P V T T L S T I T A Y L L V  
ORF 4ab

TTTGTGTGCACTTATTTTGCTTTATTCAAACCTCTTACTGCTAGAGGTGCGGTTGCTTGTTTTGTTTTAAACTATTGACACTATCTGTCI  
AAACAACAGTGAATAAAACGAAATAAGTTTGGAGAATGACGATCTCCAGCGCAACGAACAAAACAAAATTTTGATAACTGTGATAGACAGI

L F V T Y F A L F K P L T A R G R V A C F V L K L L T L S V  
ORF 4ab

ATGTGCCTTTATTGGTTCTTTTGGTATGTATCTTGACAGTTTTATAATTTTTTCTACGCTGTTGTTTCGATTCATACATGTTGGCTATT  
TACACGGAAATAACCAAGAAAAACCATACATAGAAGTGTCAAATATTAAGAAAAAGATGCGACAACAAAGCTAAGTATGTACAACCGATAA

Y V P L L V L F G M Y L D S F I I F F L R C C F D S Y M L A I  
ORF 4ab

ATGCCTATCTCTAATAAAAAATTTTTCATTTGTTTTGTTCAATGTTACTAACTATGCTTCGTTTCAGGCAAGTGTGGTATCTTGAACAATC  
TACGGATAGAGATTATTTTAAAAAGTAAACAAAACAAGTTACAATGATTTGATACGAAGCAAAGTCCGTTTACAACCATAGAAGTGTGTAG

M P I S N K N F S F V L F N V T K L C F V S G K C W Y L E Q S  
ORF 4ab

ATTTTATGAAATCGTTTTGCTGCTATTTATGGTGGTGACCACTATGTCGTTTTAGGTGGTGAAACTATTACTTTTGTTCCTTTGATGACC  
TAAAATACTTTTAGCAAAACGACGATAAATACCACCACTGGTGATACAGCAAAATCCACCACTTTGATAATGAAAACAAAGAAAAGTACTGG

F Y E N R F A A I Y G G D H Y V V L G G E T I T F V S F D D  
ORF 4ab

TTTATGTTGCTATTAGAGGTTCTTGTGAAAAGAACCTACAACCTATGCGTAAGGTTGACTTGATAATGGTGCTGTCATTTACATTTTGGC  
AAATACAACGATAATCTCCAAGAACCTTTTCTGGATGTTGAATACGCATTCCAACCTGAACATATTACCACGACAGTAAATGTAACAAACGG

L Y V A I R G S C E K N L Q L M R K V D L Y N G A V I Y I F A  
ORF 4ab

GAAGAGCCTGTTGTTGGTATAGTTTACTCCTCTCAACTATACGAAGATGTTGCTTCGATTAATTGATGACAATGGCATTGTCCTCAATTCTA  
CTTCTCGGACAACAACCATATCAAATGAGGAGAGTTGATATGCTTCTACAAGGAAGCTAATTAAGTACTGTTACCGTAACAGGAGTTAAGAT

M F L R L I D D N G I V L N S  
E

E E P V V G I V Y S S Q L Y E D V P S I N .  
ORF 4ab

TTTATGGCTCCTTGTTATGATATTTTCTTTGTGTTGGCAATGACCTTTATTAAACTGATTCAATTGTGTTTTACTTGTCAATTATTTTTT  
AAAAATACCGAGGAACAATACTATAAAAAGAAACACAACCGTTACTGGAAATAATTTGACTAAGTTAACACAAAATGAACAGTAATAAAAAAA

I L W L L V M I F F F V L A M T F I K L I Q L C F T C H Y F F  
E

AGTAGGACATTATATCAACCAGTTTATAAAATTTTCTTGCTTACCAAGATTATATGCAAATAGCACCTGTTCCAGCTGAAGTACTAAATGT  
TCATCCTGTAATATAGTTGGTCAAATATTTTAAAAGAACGAATGGTTCTAATATACGTTTATCGTGGACAAGGTCGACTTCATGATTTACA

S R T L Y Q P V Y K I F L A Y Q D Y M Q I A P V P A E V L N V  
E

CTAAACTAAACGATGTCTAATAGTAGTGTGCCTCTTTCAGAGGTTTATGTCCATTACGTAACCTGGAACCTTTAGTTGGAATTTAATTCTAAC  
GATTTGATTTGCTACAGATTATCATCACACGGAGAAAGTCTCAAATACAGGTAATGCATTGACCTTGAAATCAACCTTAATTAAGATTG

-E- M S N S S V P L S E V Y V H L R N W N F S W N L I L T  
M

AGTTTTTATAGTTGTGTTGCAGTATGGGCATTATAAGTATAGCAGACTTCTTTATGGTTTAAAGATGTCTGTTTTATGGTGTTTATGGCCAC  
TCAAAAATATCAACACAACGTCATACCGTAATATTCATATCGTCTGAAGAAATACCAAATTTCTACAGACAAAATACCACAAAATACCGGTG

V F I V V L Q Y G H Y K Y S R L L Y G L K M S V L W C L W P  
M

TTGTTCTAGCTTTGTCTATTTTTGACTGTTTTGTCAATTTAATGTGGACTGGGTCTTTTTTGGTTTTAGTATTCTTATGTCTATTATTACA  
AACAGATCGAAACAGATAAAAACTGACAAAACAGTTAAATTACACCTGACCCAGAAAAAACCAAATCATAAGAATACAGATAATAATGT

L V L A L S I F D C F V N F N V D W V F F G F S I L M S I I T  
M

CTTTGTTTATGGGTTATGTATTTGTAAATAGTTTCAGACTTTGGCGCCGTGTTAAACTTTTTGGGCTTTTAACTCCTGAACTAATGCAAT  
GAAACAAATACCCAATACATAAAACAATTATCAAAGTCTGAAACCGCGGCACAATTTGAAAAACCGGAAATTAGGACTTTGATTACGTTA

L C L W V M Y F V N S F R L W R R V K T F W A F N P E T N A I  
M

CATCTCTCTCCAGGTTTATGGACATAATTATTACTTACCGGTGATGGCTGCACCTACAGGTGTTACATTAACACTTCTTAGTGGTGTACTTC  
GTAGAGAGAGGTCAAATACCTGTATTAATAATGAATGGCCACTACCGACGTGGATGTCCACAATGTAATTGTGAAGAATCACCACATGAAG

I S L Q V Y G H N Y Y L P V M A A P T G V T L T L L S G V L  
M

TTGTTGATGGCCATAAGATTGCTACTCGTGTTCAGTGGGTCAAGTTCCTAAATATGTAATAGTTGCTACACCTAGTACCACAATTGTTGT  
AACAACTACCGGTATTCTAACGATGAGCACAAGTTCACCCAGTCAACGGATTTATACATTATCAACGATGTGGATCATGGTGTAAACAACA

L V D G H K I A T R V Q V G Q L P K Y V I V A T P S T T I V C  
M

GACCGTGTGGTGCCTCTGTTAATGAAACAAGCCAGACTGGTTGGGCATTCTACGTCCGTGCTAAACATGGTGATTTTTCTGGTGTTCCTC  
CTGGCACAACCGAGCGAGACAATTACTTTGTTTCGGTCTGACCAACCGTAAGATGCAGGCACGATTTGTACCACTAAAAAGACCACAACGGAG

D R V G R S V N E T S Q T G W A F Y V R A K H G D F S G V A S  
M



M A S V N W A D D R A  
N

A R K K F P P P S F Y M P L L V S S D K A P Y R V I P R N L

P I G K G N K D E Q I G Y W N V Q E R W R M R R G Q R V D L

P P K V H F Y Y L G T G P H K D L K F R Q R S D G V V W V A K

E G A K T V N T S L G N R K R N Q K P L E P K F S I A L P P E

L S V V E F E D R S N N S S R N A S S R S S T R N N S R D S S

R S T S R Q Q S R T R S D S N N S S S D L V A A V T L A L K N

L G F D N Q S K S P S S S G T S T P K K P N K P L S Q P R A D

K P S Q L K K P R W K R V P T R E E N V I O C F G P R D F N

ACAATATGGGGGATTGAGATCTTGTTCAGAATGGTGTGATGCCAAGGGTTTTCCACAGCTTGCTGAATTGATTCCTAATCAGGCTGCGTTA  
TGTTATACCCCTAAGTCTAGAACAAGTCTTACCACAACCTACGGTTCCTCAAAAGGTGTGCAACGACTTAAGGATTAGTCCGACGCAAT

H N M G D S D L V Q N G V D A K G F P Q L A E L I P N Q A A L  
N

TTCTTTGATAGTGAGGTTAGCACTGATGAAGTGGGTGATAATGTTTCAGATTACCTACACCTACAAAATGCTTGTAGCTAAGGATAATAAGAA  
AAGAACTATCACTCCAATCGTGACTACTTCACCCACTATTACAAGTCTAATGGATGTGGATGTTTTACGAACATCGATTCTATTATTCTT

F F D S E V S T D E V G D N V Q I T Y T Y K M L V A K D N K N  
N

CCTTCCTAAGTTCATTGAGCAGATTAGTGCTTTTACTAAACCCAGTTCTATCAAAGAAATGCAGTCACAATCATCTCATGTTGCTCAGAACA  
GGAAGGATTCAAGTAACTCGTCTAATCACGAAAATGATTTGGGTCAAGATAGTTTCTTTACGTCAGTGTTAGTAGAGTACAACGAGTCTTGT

L P K F I E Q I S A F T K P S S I K E M Q S Q S S H V A Q N  
N

CAGTACTTAATGCTTCTATTCCAGAATCTAAACCATTGGCTGATGATGATTGAGCCATTATAGAAATTGTCAACGAGGTTTTGCATTAAATT  
GTCATGAATTACGAAGATAAGGTCTTAGATTTGGTAACCGACTACTACTAAGTCGGTAATATCTTTAACAGTTGCTCCAAAACGTAATTTAA

T V L N A S I P E S K P L A D D D S A I I E I V N E V L H  
N

GTTTTGTAATTCAGTTGAATGTTTATTATTATTAGTTGCAACCCCATGCGTTTAGCGCATGATAAGGGTTTAGTCTTACACACAATGGTAG  
CAAAACATTAAGGTCAACTTACAAATAATAAATCAACGTTGGGGTACGCAATCGCGTACTATTCCCAAATCAGAATGTGTGTTACCATC

3'UTR

GCCAGTGATAGTAAAGTGTAAGTAATTTGCTATCATATTAACATGTCTAGAGGAAAGTCAGAACTTTTTCTGTTTGTGTTGTTGGAGTACTT  
CGGTCACTATCATTTACATTCATTAAACGATAGTATAATTGTACAGATCTCCTTTCAGTCTTGAAAAAGACAAACACAACACCTCATGAA

3'UTR

AAAGATCGCATAGGCGCGCCAACAATGGAAGAGCCAACAACATATCTAAAAATGTTTTGTCTGGTACTTGTTAATGATATTGTTTTGATAT  
TTTCTAGCGTATCCGCGCGGTTGTTACCTTCTCGGTTGTTGTATAGATTTTTACAAAACAGACCATGAACAATTACTATAACAAAACTATA

3'UTR

GGATACACAAAAAAAAAAAAAAAAA  
CCTATGTGTTTTTTTTTTTTTTTTT

27532

3'UTR

**Figure 2**

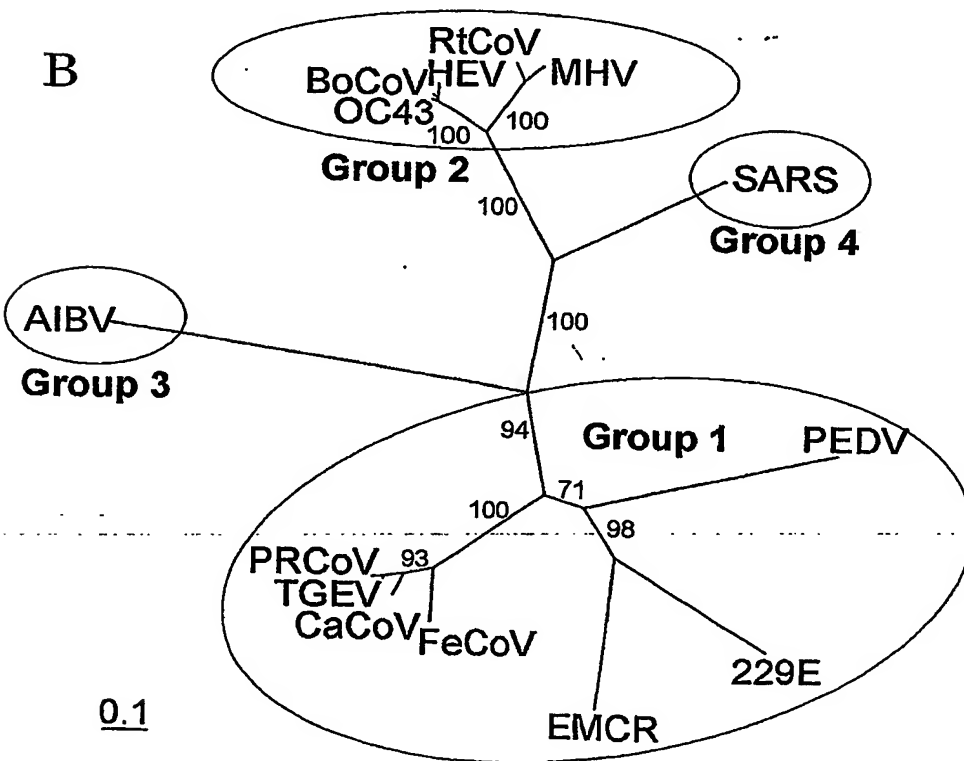
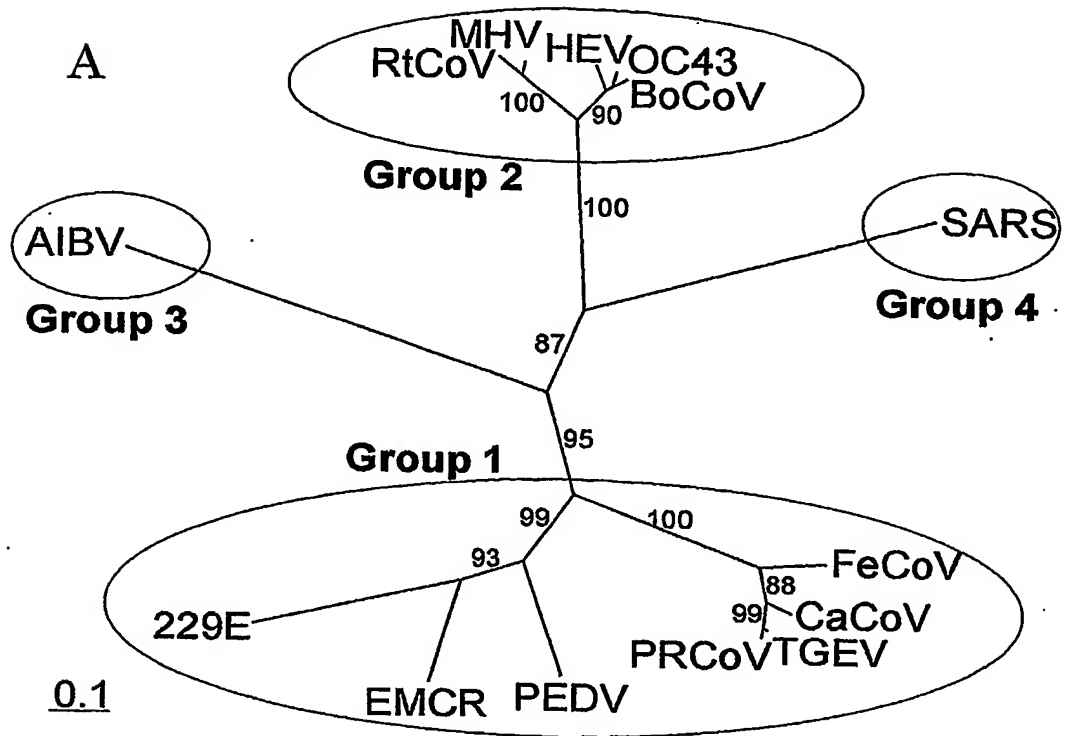


Figure 3a: Putative Orf 1a Amino acid identity matrix

Seq->	EMCR	229E	PEDV	TGEV	OC43	BoCoV	MHV	AIPV	SARS
EMCR	1.000	0.566	0.491	0.371	0.211	0.213	0.215	0.185	0.194
229E	---	1.000	0.475	0.379	0.211	0.211	0.209	0.191	0.194
PEDV	---	---	1.000	0.366	0.208	0.208	0.204	0.183	0.192
TGEV	---	---	---	1.000	0.204	0.204	0.199	0.182	0.186
OC43	---	---	---	---	1.000	0.964	0.656	0.182	0.255
BoCoV	---	---	---	---	---	1.000	0.659	0.186	0.255
MHV	---	---	---	---	---	---	1.000	0.178	0.253
AIPV	---	---	---	---	---	---	---	1.000	0.178
SARS	---	---	---	---	---	---	---	---	1.000

Figure 3b: Putative Orf 1b Amino acid identity matrix

Seq->	EMCR	229E	PEDV	TGEV	BoCoV	OC43	MHV	AIPV	SARS
EMCR	1.000	0.815	0.778	0.711	0.504	0.517	0.523	0.516	0.550
229E	---	1.000	0.765	0.720	0.504	0.520	0.515	0.515	0.546
PEDV	---	---	1.000	0.728	0.522	0.538	0.531	0.523	0.552
TGEV	---	---	---	1.000	0.517	0.533	0.529	0.520	0.551
BoCoV	---	---	---	---	1.000	0.953	0.832	0.519	0.591
OC43	---	---	---	---	---	1.000	0.857	0.534	0.611
MHV	---	---	---	---	---	---	1.000	0.532	0.606
AIPV	---	---	---	---	---	---	---	1.000	0.541
SARS	---	---	---	---	---	---	---	---	1.000

Figure 3c: Putative Orf 1ab Amino acid identity matrix

Seq->	EMCR	229E	PEDV	TGEV	OV43	BoCoV	MHV	AIBV	SARS
EMCR	1.000	0.666	0.605	0.503	0.314	0.310	0.316	0.314	0.326
229E	---	1.000	0.592	0.510	0.320	0.314	0.316	0.311	0.326
PEDV	---	---	1.000	0.509	0.326	0.320	0.323	0.313	0.328
TGEV	---	---	---	1.000	0.317	0.311	0.313	0.311	0.320
OV43	---	---	---	---	1.000	0.961	0.734	0.312	0.408
BoCoV	---	---	---	---	---	1.000	0.725	0.307	0.400
MHV	---	---	---	---	---	---	1.000	0.309	0.404
AIBV	---	---	---	---	---	---	---	1.000	0.312
SARS	---	---	---	---	---	---	---	---	1.000

Figure 3d: Putative Spike protein Amino acid identity matrix

Seq->	EMCR	229E	PEDV	TGEV	CaCoV	FeCoV	Por R	OC43	BoCoV	MHV	Rat C	PHEV	AIBV	SARS
EMCR	1.000	0.547	0.442	0.387	0.386	0.394	0.393	0.205	0.206	0.196	0.198	0.203	0.202	0.167
229E	---	1.000	0.412	0.383	0.381	0.383	0.445	0.178	0.185	0.178	0.174	0.179	0.230	0.176
PEDV	---	---	1.000	0.412	0.412	0.415	0.403	0.193	0.188	0.189	0.189	0.194	0.198	0.179
TGEV	---	---	---	1.000	0.787	0.802	0.812	0.196	0.195	0.199	0.202	0.197	0.188	0.171
CaCoV	---	---	---	---	1.000	0.911	0.743	0.196	0.195	0.202	0.202	0.194	0.184	0.167
FeCoV	---	---	---	---	---	1.000	0.758	0.195	0.193	0.197	0.200	0.196	0.188	0.167

5

10

15

20

25

30

35

40

45

[illegible]

**Figure 3e: Putative Orf E Amino acid identity matrix**

Seq→	EMCR	229E	PDV	TGEV
EMCR	1.000	0.467	0.415	0.292
229E	---	1.000	0.532	0.243
PDV	---	---	1.000	0.280
TGEV	---	---	---	1.000
CaCoV	---	---	---	---
FeCoV	---	---	---	---
Por R	---	---	---	---
OC43	---	---	---	---
BoCoV	---	---	---	---
PPHVEV	---	---	---	---
MHV	---	---	---	---
Rat C	---	---	---	---
AIBV	---	---	---	---
ARS	---	---	---	---

**Figure 3f: Matrix amino acid identity**

Seq->	EMCR	229E	229E	229E	PDGV	TGSV
EMCR	1.000	0.615	0.650	0.441		
229E	---	1.000	0.557	0.380		
PDGV	---	---	1.000	0.460		
TGSV	---	---	---	1.000		
CaCoV	---	---	---	---	---	---
FeCoV	---	---	---	---	---	---
PRCoV	---	---	---	---	---	---
OC43	---	---	---	---	---	---
HEV	---	---	---	---	---	---
LoCoV	---	---	---	---	---	---
HHV	---	---	---	---	---	---
atSA	---	---	---	---	---	---
IBV	---	---	---	---	---	---
ARS	---	---	---	---	---	---

**Figure 3g Nucleoprotein amino acid identity**

eq->	EMCR	229E	PEDV	TGEV
WCR	1.000	0,447	0,363	0,331
29E	---	1.000	0,345	0,333
EDV	---	---	1.000	0,279
GEV	---	---	---	1.000

FeCoV	---	---	---	1.000	0.756	0.763	0.212	0.213	0.192	0.195	0.197	0.217	0.185
PRCoV	---	---	---	---	1.000	0.879	0.220	0.226	0.205	0.207	0.205	0.231	0.192
CaCoV	---	---	---	---	---	---	0.216	0.220	0.199	0.201	0.199	0.217	0.199
RSDAC	---	---	---	---	---	---	1.000	0.894	0.693	0.697	0.697	0.293	0.208
MHV	---	---	---	---	---	---	---	1.000	0.680	0.684	0.682	0.290	0.216
PHEV	---	---	---	---	---	---	---	---	1.000	-0.948	0.953	0.270	0.199
OC43	---	---	---	---	---	---	---	---	---	1.000	0.973	0.272	0.203
BoCoV	---	---	---	---	---	---	---	---	---	---	1.000	0.274	0.203
SARS	---	---	---	---	---	---	---	---	---	---	---	1.000	0.206
AIBV	---	---	---	---	---	---	---	---	---	---	---	---	1.000

Figure 3h Matrix nucleotide identity

Seq->	SARS	EMCR	229E	PEDV	TGEV	CaCoV	FeCoV	PRCoV	OC43	PHEV	BoCoV	MHV	RatSA	AIBV
SARS	1.000	0.286	0.281	0.303	0.254	0.243	0.258	0.262	0.386	0.400	0.391	0.382	0.369	0.262
EMCR	---	1.000	0.615	0.650	0.441	0.429	0.441	0.437	0.317	0.317	0.317	0.303	0.303	0.239
229E	---	---	1.000	0.557	0.380	0.365	0.376	0.380	0.321	0.313	0.313	0.303	0.320	0.269
PEDV	---	---	---	1.000	0.460	0.452	0.425	0.460	0.351	0.360	0.364	0.358	0.363	0.234
TGEV	---	---	---	---	1.000	0.878	0.836	0.958	0.330	0.334	0.346	0.335	0.332	0.208
CaCoV	---	---	---	---	---	1.000	0.835	0.878	0.311	0.315	0.326	0.319	0.304	0.192
FeCoV	---	---	---	---	---	---	1.000	0.851	0.296	0.307	0.315	0.300	0.292	0.192
PRCoV	---	---	---	---	---	---	---	1.000	0.330	0.334	0.346	0.335	0.332	0.215
OC43	---	---	---	---	---	---	---	---	1.000	0.934	0.947	0.848	0.818	0.270
PHEV	---	---	---	---	---	---	---	---	---	1.000	0.943	0.848	0.818	0.270
BoCoV	---	---	---	---	---	---	---	---	---	---	1.000	0.870	0.839	0.278
MHV	---	---	---	---	---	---	---	---	---	---	---	1.000	0.938	0.271
RatSA	---	---	---	---	---	---	---	---	---	---	---	---	1.000	0.275
AIBV	---	---	---	---	---	---	---	---	---	---	---	---	---	1.000

Figure 3i Matrix nucleotide identity

Seq->	EMCR	229E	PEDV	TGEV	FeCoV	PRCoV	CaCoV	RSDAC	MHV	PHEV	OC43	BoCoV	SARS	AIBV
EMCR	1.000	0.447	0.358	0.336	0.326	0.334	0.344	0.188	0.189	0.179	0.183	0.183	0.210	0.173
229E	---	1.000	0.336	0.335	0.304	0.328	0.333	0.196	0.204	0.187	0.190	0.188	0.199	0.173
PEDV	---	---	1.000	0.277	0.244	0.272	0.270	0.163	0.168	0.160	0.160	0.158	0.184	0.178
TGEV	---	---	---	1.000	0.761	0.963	0.897	0.220	0.223	0.200	0.202	0.200	0.232	0.192
FeCoV	---	---	---	---	1.000	0.756	0.763	0.209	0.212	0.185	0.187	0.189	0.218	0.185
PRCoV	---	---	---	---	---	1.000	0.879	0.220	0.228	0.202	0.204	0.202	0.230	0.192
CaCoV	---	---	---	---	---	---	1.000	0.215	0.221	0.196	0.198	0.196	0.216	0.196
RSDAC	---	---	---	---	---	---	---	1.000	0.894	0.693	0.697	0.697	0.285	0.200
MHV	---	---	---	---	---	---	---	---	1.000	0.680	0.684	0.682	0.282	0.208
PHEV	---	---	---	---	---	---	---	---	---	1.000	0.948	0.953	0.261	0.195
OC43	---	---	---	---	---	---	---	---	---	---	1.000	0.973	0.261	0.197
BoCoV	---	---	---	---	---	---	---	---	---	---	---	1.000	0.266	0.197
SARS	---	---	---	---	---	---	---	---	---	---	---	---	1.000	0.211
AIBV	---	---	---	---	---	---	---	---	---	---	---	---	---	1.000

## Figure 4 Alignments

a. 5' untranslated region (Genomic sequence) aligned with human coronavi:  
229E

5		..... ..... ..... ..... ..... ..... ..... .....
		5 15 25 35 45 55
10	EMCR5'UTR 229E5'UTR	----- ACTTAAGTAC CTTATCTATC TACAGATAGA AAAGTTGCTT -TTTAGACTT TGTGTCTACT
		..... ..... ..... ..... ..... ..... ..... .....
		65 75 85 95 105 115
15	EMCR5'UTR 229E5'UTR	CCTCTCAACT AAACGAAATT TTT-CTAGTG CTGTCAATTG TTATG--GCA GTCCTAGTGT TTTCTCAACT AAACGAAATT TTTGCTATGG CCGGCATCTT TGATGCTGGA GTCGTAGTGT
		..... ..... ..... ..... ..... ..... ..... .....
		125 135 145 155 165 175
20	EMCR5'UTR 229E5'UTR	AATTGAAATT TCGTCAAGTT TGTA-ACCTG GTTAGGCAAG TGTTGTATTT TCTGTGTCTA AATTGAAATT TCATTGGGT TGCAACAGTT TGGAGCAAG TGCTGTGTGT CCTA-GTCTA
		..... ..... ..... ..... ..... ..... ..... .....
		185 195 205 215 225 235
25	EMCR5'UTR 229E5'UTR	AGCACTGGTG GTTCTGTG-C ACTAGTGCAC AC-ATTGATA CTTAAGT-GG TGTTCTGTCA AGGGTTTCGT GTTCCGTCTAC GAGATTCCAT TCTACAAACG CCTTACTCGA GGTCCGTCT
		..... ..... ..... ..... ..... ..... ..... .....
		245 255 265 275 285
30	EMCR5'UTR 229E5'UTR	CTGCTTATTG TGAAGCAAC GTTCTGTCTG TGTGGAAACC AATACTGCT AACC CGTGTTTGTG TGAAGCAAA GTTCTGTCTT TGTGGAAACC AGTAACTGTT CCTA

35 b. Putative Orf 1a

		..... ..... ..... ..... ..... ..... ..... .....
		5 15 25 35 45 55
40	EMCR 229E PEDV TGEV OC43 BoCoV MHV AIPV SARS CoV	----- ---MFYNQVT LAVASDSEIS GFGFAIPSPA VRAYSEAAQ GFOACRFVAF ---MACNRVT LAVASDSEIS ANGCSSTIAQA VRRYSEAAAS GFRACRFVSL ---MASNHVT LAFANDAIS AFGFCTASEA VSYSEAAAS GFMQCRFVSL ---MSSKQFV ILVNYDYQVN VPSLPIR-DV LQEIKYCYRN GFEGYVVFPE MSKINKYGLE LHWAPFPPWM FEDAEKLDN PSSSEVDMIC STTAQKLETG GICPENHVMV MSKINKYGLE LHWAPFPPWM FEDAEKLDN PSSSEVDIVC STTAQKLETG GICPENHVMV MAKMGKYGLG FKWAPEFPWM LFNASEKLGS PERSEEDGFC PSAAQEPKTK GKTILNHVRV ---MASSLKQ GVSPKPRDVI LVSKDIPEQL CDALFFYTSH NPKDYADAF ---MESLVLG VNEKTHVQLS LPVLQVRDVL VRGFGDSVEE ALSEAREHLK
		..... ..... ..... ..... ..... ..... ..... .....
		65 75 85 95 105 115
50	EMCR 229E PEDV TGEV OC43 BoCoV MHV AIPV SARS CoV	GLQDCVTGIN DDD-YVIALT GTNQLCAKIL LPSDRPLNLR GWLIFSNSNY VLQDFDVVFG DLQDCIVGIA DDT-YVMGLH GNQTLFCNIM KFSDRPFMLH GWLVFSNSNY LLEEDFVVFG DLADTVEGLL PED-YVMVVI GTTKLSAYVD TFGSRPRNIC GWLLFSNCNY FLELELTFG YCRDLVDCDR KDH-YVIGVL GNGVSDLPV LITEPSVMLQ GFIVRANCNG VLEEDFDLKIA DCRRLKQEC CVQSSLIREI VMNASPYDLE VLLQDALQSR EAVLVTPPLG MSLEACYVRG DCRRLKQEC CVQSSLIREI VMNTRPYDLE VLLQDALQSR EAVLVTPPLG MSLEACYVRG DCSRLPALEC CVQSAIIRDI FVDEDPLNVE ASTMMALQFG SAVLVKPSKR LSIQAWAKLG VROKFDRLQ TGKQKFETV CGLFLKGVVD KITPG----- VPAKVLKATS KLADLEDIFG NGTCGLVELE KGVLPQLEQP YVFIKRS DAL STNHGHKVVE LVAEMDGIQY GRSGITLGLV
		..... ..... ..... ..... ..... ..... ..... .....
		125 135 145 155 165 175
60	EMCR 229E PEDV TGEV OC43 BoCoV MHV AIPV SARS CoV	---HGAGSVVF VDKYMGFDG KPVLPKNMWE FRDYENDNTD S-IVIGGVTY QLAWDVIRKD K-RGGGNVTY TDQYLCGADG KPVMSIDLWQ FVDHFGENE E-IIIINGHTY VCAWLTKRKP ---RRGGNIVP VDQYMGADG KPVLESEWE YTDFFADSED GQLNIAGITY VKAWIVERSD ---RTGRGATY VDQYMGADG KPVIEG---D FKDYFGDED -IIEFEGEEY HCAWTTVRDE C-NPKGWTMG LFRRRSVCNT GRCTVNKHVA YQLYMIDPAG ---VCLGAGQ FVGWVIPLAF C-NPENGWTMG LFRRRSVCNT GRCAVNKHVA YQLYMIDPAG ---VCFGAGQ FVGWVIPLAF V-LPKTPAMG LFKRFCLCNT RECVCDAHVA FQLETVQPDG ---VCLGNR FIGWFVPVTA VSPLARKYRE LLKTACQWSL TVEALDVRAQ TLDEIFDPT- EILWLQVAAK VPHVGETPIA YRNVLRLKNG NKGAGGHSYG IDLKSVDLGD ELGTDPIEDY EQNWNTKHGS
		..... ..... ..... ..... ..... ..... ..... .....
		185 195 205 215 225 235
75	EMCR 229E PEDV TGEV OC43	LSYEQQNVLA IESIHLYG-T TGHTLKSCK LINA KPPKY- --SSKVVLSD EWNNAVYKAFG LDYKRONNLA IEEIEYVHGD ALHTLRNGSV LEMAKEVKT- --SSKVVLSD ALDKLYKVF VSYASONLTS IKSITYCS-T YHTFLDGT MKVARTPKI- --KKNVVLSE PLATYREIG KPLNQOTLET IQEIQYNL-D IPHKLFCAT RHVAPPVKK- --NSKIVLSE DYKKLYDIFG MPVQSRKFIV PWVMYLRKRG EKGAYNKDHG RGGFGH---- --VYDFEKVD AYDQVHDEPK

40/87

	BoCoV	MPVQSRKFIA	PWVMYLRKCG	EKGAYIKDYK	RGGFEH----	--VYNFKVED	AYDLVHDEPK	
	MHV	IPAYAKQWLQ	PWSILLRKGG	NKGSVTSGHF	RRAVTMP----	--VYDFNVED	ACEEVHLNPK	
	AIPV	IHVSSMAMLR	LVGEVTAKVM	DALGSNLSAL	FQIVKQ----	--QIARIFQK	ALAIFFENVNE	
	SARS CoV	GALRELTREL	NGGAVTRYVD	NNFCGPDGYP	LDCIKDFLAR	AGKSMCTLSE	QLDYIESKRG	
5		..... ..... ..... ..... ..... ..... .....	245	255	265	275	285	295
	EMCR	SPFTTNGISL	LDIIVKPVFF	NAFVKNCNGS	ENWSVGAWDG	YLSGCCGTPA	KKLCVVPGNV	
	229E	SPVMTNGSNI	LEAFTKPVFI	SALVQCTCGT	KSWSVGDWGT	FKSSCCNVIS	NKLCVVPGNV	
10	PEDV	SPFVDNGSDA	RSIIRRPVFL	HAFLVKCKGS	YHWTVGDWTS	YVSTCCGFKC	KPVLVASCAS	
	TGEV	SPFMNGDCL	SKCFDTLHFI	AATLRCPGCS	ESSGVGDWGT	FKTACCGLSG	KVKGVTLGDI	
	OC43	GKFSKKAYAL	IRGYRGVKPL	LYVDQYGCYD	TGSLADGLEA	YADKTLQEMK	ALFPTWSQEL	
	BoCoV	GKFSKKAYAL	IRGYRGVKPL	LYVDQYGCYD	TGSLADGLEA	YADKTLQEMK	ALFPTWSQEL	
	MHV	GKYSRKAYAL	LKGYRGVKSI	LFLDQYGCYD	TGRLAKGLED	YGDCTLEEMK	ELFPVWCDSL	
15	AIPV	LEQRIAAALM	AFAKCARSIT	VVVVERTLVV	KEFAGTCLAS	INGAVAKFFE	ELPNGMFGSK	
	SARS CoV	VYCCRDHEHE	IAWFTERSDK	SYEHQTPFEI	KS--AKKEDT	FKGECPKFVF	PLNSKVVKVIQ	
20		..... ..... ..... ..... ..... ..... .....	305	315	325	335	345	355
	EMCR	VEGDVIITST	DAGCGVKYVA	GLVVKHITNI	TGVSILWRVTA	VHSDGMFVAT	SSYDALLHRN	
	229E	KPGDAVITTO	QAGAGIKYFC	GMTLKFVANI	EGVSVWRVIA	LQSVDCFVAS	STFVEEEHVN	
	PEDV	MPGSVVVTRA	GAGTGKYYN	NMFLRHVADI	DGLAFWRILK	VQSKDDLACS	GKFLHHEEG	
	TGEV	KPGDAVVTSM	SAGKGVKFFA	NCVLQYAGDV	EGVSIWKVIK	TFTVDETCTV	PGFEGELN--	
	OC43	LFDDVIVAWHV	VRDP-----RY	VMRLQSAATI	R---SVAYVA	NPTEDLDCGS	VVIKEPVHVV	
25	BoCoV	PFDDTVAWHV	VRDP-----RY	VMRLQSAATI	R---SVAYVA	NPTEDLDCGS	VVIKEPVHVV	
	MHV	DNEVVVWAWHV	DRDP-----RA	VMRLQTLATI	R---SIGYVG	QPTEDLVDDG	VVVREPAHLL	
	AIPV	IFTTLAFAKE	AAVR-----	-VVENIPNAP	RGTKGFEVVG	NAKGTQVVVR	GMRNDLTLLD	
	SARS CoV	PRVEKKKTEG	FMGRIRSVYP	VASPBECNNM	HLSTLMKCNH	CDEVSWQTC	FLKATCEHCG	
30		..... ..... ..... ..... ..... ..... .....	365	375	385	395	405	415
	EMCR	SLDPFCFDVN	TLLSNQLRLA	FLGASVTEDV	KFAASTGVID	ISAGMFLGYD	DILTNNKPWF	
	229E	RMDTFCFNVR	NSVTDECRLA	MLGAEMTSNV	RRQVASGVID	ISTGWFDDVD	DIFAESKPWF	
	PEDV	FTDPCYFLND	SSLATKLKFD	ILSGKFSDEV	KQAIAGHV	VGSALVDIVD	DALG--QPWF	
35	TGEV	--DFIKPESK	SLVACSVKRA	FITGDIDDAV	HDCIITGKLD	LSTNLFNGVG	LEFKK--TPWF	
	OC43	ADDSIILRQY	NLVDIMSHFY	MEADTVVNAF	YGVALKDCGF	VMQFGYIDCE	QDSCDFKGIW	
	BoCoV	ADDSIILRQH	NLVDIMSHFY	MEADTVVNAF	YGVALKDCGF	VMQFGYIDCE	QDSCDFKGIW	
	MHV	AANAIVKRLP	RLVETMLYT--	--DSSVTEFC	YKTKLDCGCF	ITQFGYVDDC	GDACDFRGWV	
	AIPV	QKADIPVEPE	GWSAILDGH	CYVFRSGDRF	YAAPLSGNFA	LSDVHCCERV	VCLSDGVTPE	
40	SARS CoV	-TENLVIEGP	TTCGYLPTNA	VVKMPCPACQ	DPEIGPEHSV	ADYHNHSNIE	TRLR--KGG	
45		..... ..... ..... ..... ..... ..... .....	425	435	445	455	465	475
	EMCR	VRKASGLFEDA	IWDAFVAAIK	LVPTTTGGLV	RFVKSIASTV	LTVSNGVIIM	CADVPDAFQP	
	229E	VRKAEDIFGP	CWSALASALK	QLKVTGTGLV	RFVKSICNSA	VAVVGGTIQI	LASVPEKFLN	
	PEDV	IRKLGDLASA	PWEQLKAVVR	GLGLLSDEVV	LFGKRLSCAT	LSIVNGVEFF	LADVPCKLAA	
	TGEV	VQKCGALFVD	AWKVVEELCG	SLTLTYKQIY	EVVASLCTSA	FTIVNYKPTF	VVPD--NRVKD	
	OC43	PGNMIDGFAC	TTCGHVYEVG	DLIAQSSGVL	PVNPVLHTKS	AAGYGG----	-FGCKDSFTL	
50	BoCoV	PGNMIDGFAC	TTCGHVYEVG	DLIAQSSGVL	PVNPVLHTKS	AAGYGG----	-FGCKDSFTL	
	MHV	PGNMMDGFLC	PGCSKSYMPW	ELEAQSSGVI	PKGGVLFQTS	TDTVN-----	-----RESFKL	
	AIPV	INDGLILAAI	YSSFSVSSELV	TALKKGEPFK	FLGHKFVYAK	DAAVS-----	-----FTL	
	SARS CoV	TRCFGGCVFA	YVGCYNKRAY	WVPRASADIG	SGHTGITGDN	VETLN-----	-----EDLLEILS	
55		..... ..... ..... ..... ..... ..... .....	485	495	505	515	525	535
	EMCR	VYRTFTQAIC	AAFDLSLDVF	KIG---DVKE	KRLGDYVLTE	NALVRLTTEV	VRGVRDARIK	
	229E	AFDVFVTAIQ	TVEDCAVETC	TIA---GKAF	DKVFDYVLLD	NALVKLVTTK	LKGVREGRGN	
	PEDV	AVTVFVNFLN	EFFESACDCL	KVG---GKTF	NKVGSYVLF	NALVKLVKAK	ARGPRQAGIC	
	TGEV	LVQKCVKVLV	KAFDVFTQII	TIAGIEAKCF	VLGAKYLLFN	NALVKLVSVK	ILGKQKQGLE	
60	OC43	YGQTVVYFGG	CVYWSPARNI	WIP---ILKSS	VKSYSLSLVY	GVLGCKAIVK	ETNLICKALY	
	BoCoV	YGQTVVYFGG	CVYWSPARNI	WIP---ILKSS	VKSYSLSLVY	GVLGCKAIVK	ETNLICKALY	
	MHV	YGHAVVPFGS	AVYWSPYPGM	WLP---VIWSS	VKSYSADLT	GVLGCKAIVQ	ETDAICRSLY	
	AIPV	AKAATIADVL	RLFQSARVIA	EDVWS-SFTE	KSFEEFKLAY	GKVRNLEEFV	KTYVCKAQMS	
	SARS CoV	REVRNINIVG	DFHLNEEVAI	ILAS-FSAST	SAFIDTIKSL	DYKSFKTIVE	SCGNKYKVTG	
65		..... ..... ..... ..... ..... ..... .....	545	555	565	575	585	595
	EMCR	KAMFTKVVG	PTTEVKFSVI	ELATVNLRLV	DCAPVVCPPK	KIVVIAGQAF	FYSGGFYRFM	
	229E	KVKYATVVVG	STEEVKSSRV	ERSTAVLTIA	NNYSKLEDEG	YTVVIGDVAY	FVSDGYFLM	
	PEDV	EVRYTSLVVG	STTKVVSKRV	ENANVNLVVV	DEDVTLNTTG	RTVVVDGLAF	FESDGFYRHL	
	TGEV	CAFFATSLVG	ATVNVTPKRT	ETATISLNKV	DDVAVP-EG	YIVIVGDMAF	YKSGEYFMM	
	OC43	LDYVQHKCGN	LHQRELLGVS	DVWHKQLLLN	RGVYKPLEN	IDYFNMRRAK	FSLETFTVCA	
	BoCoV	LDYVQHKCGN	LHQRELLGVS	DVWHKQLLLN	RGVYKPLEN	IDYFNMRRAK	FSLETFTVCA	
	MHV	MDYVQHKCGN	LEQRAILGLD	DVYHRLQLVN	RGDYSLLLEN	VDLFVKRAAE	FACK-FATCG	
75	AIPV	IVILAALVGE	DIWHLVSQVI	YKLGVLFTKV	VDFCDKHWKG	FCVQLKRAKL	IVTETFCVLK	
	SARS CoV	KPVKGAWNIG	QQRSVLTPLC	GFPSQAAGVI	RSIFARTLDA	ANHSIPDLQR	AAVTILDGIS	
80		..... ..... ..... ..... ..... ..... .....	605	615	625	635	645	655
	EMCR	VDSTTVLNDP	VFTGELFYTI	KFSGFKLDGF	N-----H	QFVNASSATD	ATIIVELLLS	



229E	ASPNSVLTTA	VYKPLFAENV	NVMGTRPE--	-----	KFPPTVTTCEN	LESADVLFVND
PEDV	ADADVIEHP	VYKSACELKP	VFECDDIP--	D-----F	PLPVAASVAE	LCVQTDLLLK
TGEV	SSPNFVLTTN	VFKAVKVPSY	DIVYDVNDNT	KSKMIAKLGS	SFEYDGDIDA	AIVKVNELLI
OC43	DGFMPELDD	LVPRAVYLA	SGQAFCDY--	-----	ADKLCHAVVS	KSKELLDVSL
BoCoV	DGFMPELDD	LVPRAVYLA	SGQAFCDY--	-----	AGKICHAVVS	KSKELLDVSV
MHV	DGLVPLLLDG	LVPRAVYLIK	SGQAFCTSM--	-----	MVNFSEHVD	MCMDMALLFM
AIPV	GVAQHCFQLL	LDAHSYLYKS	FKKCALGR--	-----	---IHGDLF	WKGGVHKIVQ
SARS CoV	EQSLRLVDAM	VYTSDDLNTS	VIIMAYVTG-	-----	--GLVQQTSTQ	WLSNLLGTTV
10	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	665	675	685	695	705	715
EMCR	DFKTAVFVYT	CVVDGCSVIV	RRDAT-FATH	VCFKDCYSIW	EQFCIDNCGE	PWFLLTDYNAI
229E	KITEFQLDYS	IDVIDNEIIV	KPNIS-LCVP	LYVRDYVDKW	DDFCRQYSNE	SWFEDDYRAF
PEDV	NYNTPYKYTS	CVVRGDKCCI	TCTLO-FKAP	SYVEDAVN-F	VDLCTKNIGT	AGFEFYITA
TGEV	EFRQQLSCLFR	AFKDDKSIFV	EAYFKKYKMP	ACLAKHIG-L	WNIKKDSCK	RGFLNLFNHL
OC43	DSLGAIIHYL	NSKIVDLAQH	FSDFG-----	TSFVSKIVHF	FKTFTTSTAL	AFAWVLFHVL
BoCoV	DSLGAIIHYL	NSKIVDLAQH	FSDFG-----	TSFVSKIVHF	FKTFTTSTAL	AFAWVLFHVL
MHV	HDVAVATKYV	KKVTGKLAVR	FKALG-----	VAVVRKITTEW	FDLAVDTAAS	AAGWLCYQLV
AIPV	DGDEIWFDAI	DSVDVEDLGV	VQEKs---ID	FEVCDDVTLP	ENQPGHMQVI	EDDGKNYMEF
SARS CoV	EKLRFIFEWI	EAKLSAGVEF	LKDAW-----	EILKFLITGV	FDIVKGQIQV	ASDNIKDCVK
15	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	725	735	745	755	765	775
EMCR	LQSNPPQCAI	VQASESK--V	LLERFLPKCP	EILLSIDDGH	LWNLFVEKFN	FVTDWLKTLK
229E	ISVLDITDAA	VKAESK--A	FVDITVPPCP	SILKVIDGGK	IWNGVIKNVN	SVRDWLKSLK
PEDV	HEQQDLQGGF	TTCCTMSGFE	CEMPTIQCPC	AVLEEIDGGS	IWRSFITGLN	TMDWDFCKRLK
TGEV	NELEDIKETN	IQAIKN-----	-----ILCP	DPILLDLGYA	IWYNCMPGCS	DP-SVLGSVQ
OC43	HGAYIVVESD	IYFVKN-----	-IPRYASAVA	QAFQSVAKVV	LDSLRVTFFD	GLSCFKIGRR
BoCoV	HGAYIVVESD	IYFVKN-----	-IPRYASAVA	QAFQSVAKVV	LDSLRVTFFD	GLSCFKIGRR
MHV	NGLFAVANGG	ITFLSD-----	-VPELVKNFV	DKEKVFVKVL	IDSMSVSVLS	GLTVVKTASN
AIPV	RFKKDENIYY	TPMSQLG----	---AINVVCK	AGGKTVTFG-	--ETTVOEIP	PPDVVPIKVS
SARS CoV	CFIDVVNKAL	EMCIDQ-----	---VTIAG	AKRLSLNLGE	VFTAQSKGLY	RQCIRGKEQL
20	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	785	795	805	815	825	835
EMCR	LTLTSNGLLG	NCAKRFRVL	VKLLDVNGF	LETVCSSVHT	AGVCIKYYAV	NVP-YVVISG
229E	LNLTOQGLLG	TCAKRFRKWL	GILLEAYNAF	LDTVVSTVKI	GGTFFKTYAF	DKP-YIVIRD
PEDV	VSFGLDGIVV	TVARKFKRLG	ALLAEMYNTY	LSTVVENLVL	AGVSFKYYAT	SVP-KIVLGG
TGEV	LLIGNG--VK	VVCDGCKGFA	NQLSKGYNKL	CNAARNDEI	GGIPFSTFT	PTNTFIEMTD
OC43	RICLSGRKIY	EVERGLLHSS	QLPLDVYDLT	MPSQVQKAKQ	KPIYLKSGSG	DFSLADSVVE
BoCoV	RICLSGRKIY	EVERGLLHSS	QLPLDVYDLT	MPSQVQKAKQ	KPIYLKSGSG	DFSLADSVVE
MHV	RVCLAGCKVY	EVVQKRLSAY	VMPVGCNEAT	C-----	----LVGEIE	PAVVEDDDVD
AIPV	IECCGEPWNT	IFKKAYKEPI	EVDTLTVEQ	LLSVIYEKMC	DDLKLFPEAP	EPFPPFENVAL
SARS CoV	QLLMPLKAPK	EVTFLLEGDSH	DTVLTSEEVV	LKNGELEALE	TPVDSFTNGA	IVGTPVCVNG
45	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	845	855	865	875	885	895
EMCR	FVSRVIRRR	CD--VFPPCV	SCVTFFYEFL	DTCFGVSK--	--PNAIDVEH	LELKETVFE
229E	IVCKVENKTE	AEWIELEPHN	DRIKSFTTFF	SAYMPIAD--	--PTHFDIEE	VELLDAEFVE
PEDV	CFHSVKSVEA	SV--FQIPVQ	AGIEKFKVFL	NCVHPVV--	--PRVIETSF	VELEETTFKP
TGEV	AIYSVIEQK	AL-----	S-----	DADVPVVDNG	TISTADWSEP	ILLEPAEYVK
OC43	VVTSLTPCG	YS-----EPP	KVADKICIVD	NVYMAKAGDK	YYPVVVD--DH	VGLLDQAWRV
BoCoV	VVTSLTPCG	YS-----EPP	KVADKICIVD	NVYMAKAGDK	YYPVVVD--GH	VGLLDQAWRV
MHV	VVKAPLTYYQ	CC-----KPP	TSFEKICVVD	KLYMAKCGDQ	FYPVVVDNDT	IGVLDQCWR
AIPV	VDKNGKDLDC	IKS-----	-----CHLI	YRDYESD--	-----DD	IEEDEAECD
SARS CoV	LMLEIKDKKE	QY-----	-----CALS	PGLLATN--	-----NV	FRKKGGAIPK
50	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	905	915	925	935	945	955
EMCR	PKDGGQFFVS	DDYLWYVV-D	D-----IY	YPASCNGVLP	VAFTKLGGK	---ISFSDDV
229E	PGCGGILAVI	DEHVYFKK-D	G-----VY	YPSNGTNILP	VAFTKAAGGK	---VSFSDDV
PEDV	PALNGGIAIV	DGFAFYD-G	T-----LY	YPTDGNSSVP	ICFKKKGGGD	---VKFSDEV
TGEV	PKNNGNVIVI	AGYTFYKDED	E-----HF	YPYGFQKIVQ	RMYNKMGGGD	KT-VSFSEV
OC43	PCAGRRTVTFK	EQPTVKEIIS	MPKIIKVFYE	LDNDNFNTILN	TACGVFEVDD	TVDMEEFYAV
BoCoV	PCAGRCVTFK	EQPTVNEIIS	TPKTIKVFYE	LDKDFNTILN	TACGEFEVDD	TVDMEEFYAV
MHV	PCAGKKVEFN	DKPKVKEIPS	T-RKIKINFA	LDATFDSVLS	KACSEFEVDK	DVTLDLDDV
AIPV	TDSGEAEEDC	TNSECEEEDE	D-----TK	VIALIQDPAS	IKYPLPLEDD	YS-VYNGCIV
SARS CoV	GVTFGEDTVW	EVQGYKNVRI	T-----FE	LDERVDRVLN	EKCSVYTVES	GTEVTEFACV
60	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	965	975	985	995	1005	1015
EMCR	IVHDVEPTHK	VKLIFEFEDD	-VVTSLCKKS	FGKSIITYTG-	DWEGLEHVL	SAMNVIG--
229E	EVKDIEPVYR	VKLCEFEFE	-KLVDVCEKA	IGKKIKHEG-	DWDSFCKTIQ	SALSVVS--
PEDV	SVRTIDPVYK	VSLEFEFESE	-TIMAVLNKA	VGNRIKVTG-	GWDDVVEYIN	VAIEVLK--
TGEV	DVQEIAPVTR	VKLEFEFENE	-IVTGLERA	IGTRYKFTGT	TWEEFEESIS	EELDAIFDTL
OC43	VIDAIEEKLS	PCKELEGVGA	-KVSAPLQKL	EDNPLFLFD-	--EAGEEVLA	PKLYCAFTAP
BoCoV	VIDAIEEKLS	PCKELEGVGA	-KVSAPLQKL	EDNPLFLFD-	--EAGEEVLA	PKLYCAFTAP
MHV	VLDVESTLS	PCKEHDVIGT	-KVCALLNRL	AEDYVYLFD-	--EGGEEVIA	PKMYCSFSAP
AIPV	HKDALDVVNL	PSGEETFFVN	NCFEGAVKPL	PQKVVDVLG-	--DWGEAVDA	QEQLCQ--
SARS CoV	VAEAVVKTLQ	PVSDLLTN--	--MGIDLDEW	SVATFYLF-	--DAGEENFS	SRMYCSFYPP
70	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	965	975	985	995	1005	1015
EMCR	IVHDVEPTHK	VKLIFEFEDD	-VVTSLCKKS	FGKSIITYTG-	DWEGLEHVL	SAMNVIG--
229E	EVKDIEPVYR	VKLCEFEFE	-KLVDVCEKA	IGKKIKHEG-	DWDSFCKTIQ	SALSVVS--
PEDV	SVRTIDPVYK	VSLEFEFESE	-TIMAVLNKA	VGNRIKVTG-	GWDDVVEYIN	VAIEVLK--
TGEV	DVQEIAPVTR	VKLEFEFENE	-IVTGLERA	IGTRYKFTGT	TWEEFEESIS	EELDAIFDTL
OC43	VIDAIEEKLS	PCKELEGVGA	-KVSAPLQKL	EDNPLFLFD-	--EAGEEVLA	PKLYCAFTAP
BoCoV	VIDAIEEKLS	PCKELEGVGA	-KVSAPLQKL	EDNPLFLFD-	--EAGEEVLA	PKLYCAFTAP
MHV	VLDVESTLS	PCKEHDVIGT	-KVCALLNRL	AEDYVYLFD-	--EGGEEVIA	PKMYCSFSAP
AIPV	HKDALDVVNL	PSGEETFFVN	NCFEGAVKPL	PQKVVDVLG-	--DWGEAVDA	QEQLCQ--
SARS CoV	VAEAVVKTLQ	PVSDLLTN--	--MGIDLDEW	SVATFYLF-	--DAGEENFS	SRMYCSFYPP
75	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	965	975	985	995	1005	1015
EMCR	IVHDVEPTHK	VKLIFEFEDD	-VVTSLCKKS	FGKSIITYTG-	DWEGLEHVL	SAMNVIG--
229E	EVKDIEPVYR	VKLCEFEFE	-KLVDVCEKA	IGKKIKHEG-	DWDSFCKTIQ	SALSVVS--
PEDV	SVRTIDPVYK	VSLEFEFESE	-TIMAVLNKA	VGNRIKVTG-	GWDDVVEYIN	VAIEVLK--
TGEV	DVQEIAPVTR	VKLEFEFENE	-IVTGLERA	IGTRYKFTGT	TWEEFEESIS	EELDAIFDTL
OC43	VIDAIEEKLS	PCKELEGVGA	-KVSAPLQKL	EDNPLFLFD-	--EAGEEVLA	PKLYCAFTAP
BoCoV	VIDAIEEKLS	PCKELEGVGA	-KVSAPLQKL	EDNPLFLFD-	--EAGEEVLA	PKLYCAFTAP
MHV	VLDVESTLS	PCKEHDVIGT	-KVCALLNRL	AEDYVYLFD-	--EGGEEVIA	PKMYCSFSAP
AIPV	HKDALDVVNL	PSGEETFFVN	NCFEGAVKPL	PQKVVDVLG-	--DWGEAVDA	QEQLCQ--
SARS CoV	VAEAVVKTLQ	PVSDLLTN--	--MGIDLDEW	SVATFYLF-	--DAGEENFS	SRMYCSFYPP
80	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	965	975	985	995	1005	1015
EMCR	IVHDVEPTHK	VKLIFEFEDD	-VVTSLCKKS	FGKSIITYTG-	DWEGLEHVL	SAMNVIG--
229E	EVKDIEPVYR	VKLCEFEFE	-KLVDVCEKA	IGKKIKHEG-	DWDSFCKTIQ	SALSVVS--
PEDV	SVRTIDPVYK	VSLEFEFESE	-TIMAVLNKA	VGNRIKVTG-	GWDDVVEYIN	VAIEVLK--
TGEV	DVQEIAPVTR	VKLEFEFENE	-IVTGLERA	IGTRYKFTGT	TWEEFEESIS	EELDAIFDTL
OC43	VIDAIEEKLS	PCKELEGVGA	-KVSAPLQKL	EDNPLFLFD-	--EAGEEVLA	PKLYCAFTAP
BoCoV	VIDAIEEKLS	PCKELEGVGA	-KVSAPLQKL	EDNPLFLFD-	--EAGEEVLA	PKLYCAFTAP
MHV	VLDVESTLS	PCKEHDVIGT	-KVCALLNRL	AEDYVYLFD-	--EGGEEVIA	PKMYCSFSAP
AIPV	HKDALDVVNL	PSGEETFFVN	NCFEGAVKPL	PQKVVDVLG-	--DWGEAVDA	QEQLCQ--
SARS CoV	VAEAVVKTLQ	PVSDLLTN--	--MGIDLDEW	SVATFYLF-	--DAGEENFS	SRMYCSFYPP

		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....
		1025 1035 1045 1055 1065 1075
5	EMCR	--QHKLPPQF YIYDEEGGYD VSKP--VMIS QWPISSDDSDG CVVEASTDFH Q--LESVREE
	229E	--CYVNLPTY YIYDEEGGND LSLP--VMIS EWPLSVQQAQ QEATLPDIAE D--VVDQVEE
	PEDV	--DHVEVPKY YIYDEEGGTD PNLK--VMVS QWPLNDDTIS QDLLDVEVVT DAPIDSEGDE
	TGEV	ANQGVLEGY FIYDTCGGFD IKNPDGIMIS QYDINITADE KSEVSASSE EE--VESVEED
	OC43	EDDDFLEESD VEEDDVEGEE TDLTVTSAGQ PCVASEQEE SEVLEDTLDD GPSVETSDSQ
	BoCoV	EDDDFLEESG VEEDDVEGEE TDLTVTSAGE PCVASEQEE SEILEDTLDD GPCVETSDSQ
10	MHV	DDEDCAADV VDADENQGGD ADDSAALVTD TQEDDGVAKG QVGVAESDAR LDQVEAFDIE
	AIPV	--EPLQHTFE EPVENSTGSS KTMTEQVVVE DQELPVVEQD QDVVVYPTPD LEVAKETAEE
	SARS CoV	DEEEEDDAEC EEEEIDETCE HEYGTEDDYQ GLPLEFGASA ETVRVEEEEE EDWLDLDTTEQ
		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....
		1085 1095 1105 1115 1125 1135
15	EMCR	VD----- IIE QPFGEVEHAL SIRO-----
	229E	VNS----- IFD IETVDVKHDV S-----
	PEDV	VDSSAPEKVA -----D VANSEPGDDG LPVAPETNVE SEVEEVAATL SFIKDTPSTV
	TGEV	PENEIVEASE GAEGTSSQEE VETVEVADIT STEEDVDIVE VSAKDDFWAA AVDVQAEQF
20	OC43	VEEDVEMS-- -----DFVDL ESVIQD----- --YENVCFEF YTT-----
	BoCoV	VEEDVQMS-- -----DFGDL ESVIQD----- --YENVCFEF YTT-----
	MHV	KVEDPILN-- -----ELSAE LNAPADK-- --TYEDVLAFD AIYSEALSAF YAVP-----
	AIPV	VD-----
	SARS CoV	SEIEPEP-----E PTPEEPVNQF TG-----
25		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....
		1145 1155 1165 1175 1185 1195
30	EMCR	---PFSFSFR DELGVRVLDQ SDNNCWISTT LIQLQLTKLL DDSIEMQLFK VGVKVDISVQK
	229E	---PFEMPFE ELNGLKILKQ LDNNCWVNSV MLQIQLTGIL DGDYAMQFFK MGRVAKMIER
	PEDV	TKDFAFADFV SYGGLKVLRL SHNNCWVSTT LVQLQLLGIV DDP--AMELFS AGRVGPMVRK
	TGEV	NPSLPPFKTT NLNGKILKQ GDNNCWINAC CYLOAQEDFF NNE--AWKEFK KGDVMDVFNL
	OC43	-----EPEFV KVLGLYVPA TRNNCWLRV LAVMQLKPCQ FKD--KNLQD LWVLYKQQYS
	BoCoV	-----EPEFV KVLGLYVPA TRNNCWLRV LAVMQLKPCQ FKD--KNLQD LWVLYKQQYS
	MHV	-----GDETHF KVCGFYSPAI ERTNCWLRSV LIVMQSLPLE FKD--LEMOK LWLSYKSSYN
35	AIPV	----- EFILIFAVPK EEVVSQRDGA QIKQEPQV KPQ--REKKA KKFVKVPATC
	SARS CoV	-----YLKLTN NVAIKCVDIV KEAQSANPMV IVNAANIHLK HGGGVAGALN KATNGAMQKE
		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....
		1205 1215 1225 1235 1245 1255
40	EMCR	CYELSHLISG SLGDSGKLLS ELLKDKYTCS ITFEMSCDCG KKFDEQVGC LFWIMPYTKLF
	229E	CYTAEQCIRG AMGDVGLCMY RLLKDLHTGF MVMDYKCSCT SGRLEESGAV LFCTPTTKAF
	PEDV	CYESQKAILG SLGDVSACLE SLTKDLHTLK ITCSVVCSCG TGERIYEGCA FRMTPTLEPF
	TGEV	CYAATTLARG HSGDAEYLLE LMLNDYSTAK IVLAAKCGCG EKEIVLERAV FKLTPLEKSF
	OC43	QLFVDTLVNK IPANIVLPQG GYVADFAYWF LTLCDWQCVA YWKCIKCDLA LKLKGLDAMF
45	BoCoV	QLFVDTLVNK IPANIVLPQG GYVADFAYWF LTLCDWQCVA YWKCIKCDLA LKLKGLDAMF
	MHV	KEFVDKLVKS VKPSIILPQG GYVADFAYWF LSQCSFKAYA NWRCLKCDMD LKLQGLDAMF
	AIPV	EKPKFLEYKT CVGDLTVVIA KALDEFKEFC IVNAANEHMT HSGGVAKAIA DFCGLDFVEY
	SARS CoV	SDDYIKLNGP LTVGGSCLLS GHNLAKKCLH VVGPNLNAGE DIQLLKAAYE NFNSQDILLA
		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....
		1265 1275 1285 1295 1305 1315
50	EMCR	QKGECCICHK MQTYKLVSMT GTGVFVQD-- PAPIDIDAFV VRPICSSVYL GVKGSQGHYQT
	229E	PYGTCLNCNA PRMCTIRQLQ GTIIFVQOK-- PEPVNPVSFV VKPVCSSIFR GAVSCGHYQT
	PEDV	PYGACAQCAQ VLMHTFKSIV GTGIFCRD-- TTALSLSLSV VKPLCAAIFI GK--DSGHYVT
55	TGEV	NYGVCGDCMQ VNTCRFLSVE GSGVFVHDIL SKQTPAEAFV VKPVMHAVYT GTQNGHYMV
	OC43	FYGDVVSCHC KCGESMVLID VDVPFTHAFA LKDKLFCAFI TKRIVYKAAC VVDVNDSHSM
	BoCoV	FYGDVVSCHC KCGESMVLID VDVPFTHAFA LKDKLFCAFI TKRSVYKAAC VVDVNDSHSM
	MHV	FYGDVVSCHC KCGTGMTLIS ADIPYTLHFG LRDDKECAFY TPRKVFRAAC VVDVNDCHSM
	AIPV	CEDYVKKHGP QORLVTPSFV KGIQCVNNV GPRHGDNNLH EKLVAAYKNV LVDGVVNVYV
60	SARS CoV	PLLSAGIFGA KPLQSLQVCV QTVRTQVYIA VNDKALYEQV VMDYLDNLKP RVEAPKQEEP
		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....
		1325 1335 1345 1355 1365 1375
65	EMCR	NLYSFDKID GFGVFDIK-- -----NSSV NTVCFVDVDF HS--VEIEAGE
	229E	NIYSQNLCDV GFGVNKIQP-- -----WTNDAL NTICIKDADY NAKVEISVTP
	PEDV	NFYDAAMAID GYGRHQIK-- -----YDTL NTICVKDWNW TABLVPAVDS
	TGEV	DDIEHGYCDV GMGIKPLKKR CYTSTLFINA NVMTRAKEPK QEFKVEKVEQ QPIVEENKSS
	OC43	AVVDG--KQID DHRITSIT-- -----SDK FDFIIGHGMS FSMTTFEIAQ
	BoCoV	AVVDG--KQID DHRITSIT-- -----SDK FDFIIGHGMS FSMTTFEIAQ
70	MHV	AVVDG--KQID GKVVTKFN-- -----GDK YDFMVGHGMA FSMSAFEIAQ
	AIPV	PVLSLGIFGV DFKMSIDAMR -----EA FEGCTIRVIL FSLSQEHIDY
	SARS CoV	FNTEDSKTEE KSVVQKPVVD KP----- --KIKACIDE VTTTLEETKF LTNKLILLFAD
		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....
		1385 1395 1405 1415 1425 1435
75	EMCR	VK----- PFAVYKNVKF YLGDISHLVN CVSFDFVVNA ANENLMHGGG
	229E	IKNTVDTPPK EEFVVKELN AFLVHDNVAF YQGDVDTVVN GVDFDFIVNA ANENLAHGGG
	PEDV	VVEP----- VVK PFYKKNVDF YQGDVDTLVK --LPCDFVVNA ANEKLSHGGG
	TGEV	IEKEEIQSPK ND-----DLIL PFYKAGKLSF YQGALDVLIN FLEPDVIVNA ANGDLKHMGG
80	OC43	LYG----- --SCITPNVCF VKGDIKIVSK LVKAEEVVNP ANGHMAHGGG

	BoCoV	LYG-----	-----	-SCITPNVCF	VKGDIKVKSK	RVKAEVVVNP	ANGHMAHGGG
	MHV	LYG-----	-----	-SCITPNVCF	VKGDIKVKLR	RVGAEVIVNP	ANGRMAHGGG
	AIPV	FD-----	-----	-----VTC	KQKTYLTED	GVKYRSIVLK	PGDSLQGFQ
5	SARS CoV	INGKLYHD--	-----SQ	NMLRGEDMSF	LEKDAPYVMG	DVITSGDITC	VVIPSCKKAGG
		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		1445	1455	1465	1475	1485	1495
	EMCR	VARAIDILTE	GQLQSLSKDY	ISSNGPLKVG	AGVMLE--CE	KFNVFNVVGP	RTG---KHEH
10	229E	LAKALDVYTK	GKLQRLSKEH	IGLAGKVKVG	TGVMVE--CD	SLRIFNVVGP	RKG---KHER
	PEDV	IAKAIDVYTK	GMLQKCSNDY	IKAHGPIKVG	RGVMLE--AL	GLKVFNVVGP	RKG---KHAP
	TGEV	VARAIDVFTG	GKLTERSXDY	LKKNKSIAPG	NAVFFENVIE	HLSVLNAVGP	RNGD---SRVE
	OC43	VAKAIAVAAG	QQFVKETTDM	VKSKGVCATG	DCYVSTGGKL	CKTVLNVVGP	DARTQKGQSY
	BoCoV	VAKAIAVAAG	QQFVKETTDM	VKSKGVCATG	DCYVSTGGKL	CKTVLNVVGP	DARTQKGQSY
15	MHV	VAGAIKAAAG	KSFIKETADM	VKNQGVQCVG	ECYESTGGNL	CKTVLNVVGP	DARTQKGQSY
	AIPV	VYAKNKIVFT	ADDVEDKEIL	YVPTTDKSIL	EYYGLD---A	QKYVIYLOTL	AQKWNVQYRD
	SARS CoV	TTEMLSRALK	KVPVDEYITT	YPQGGCAGYT	LEEAKTALKK	CKSAFYVLPS	EAPNAKEEIL
		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		1505	1515	1525	1535	1545	1555
20	EMCR	SLLVEAYNSI	LF---ENGIP	LMPLLSGIF	GVRIENSLKA	LFSCDINKPL	QVFVYSSNEE
	229E	DLLIKAYNTI	NN---EQGTP	LTPLISGIF	GIKLETSLEV	LLDVCNTKEV	KVFVYTDTEV
	PEDV	ELLVKAYKSV	FA---NSGVA	LTPLISVGIF	SVPLEESLSA	FLACVGDHRC	KCFYGDKER
	TGEV	AKLCNVYKAI	AK---CEGKI	LTPLISVGIF	NVRLETSLOC	LLKTVNDRGL	NVFVYTDQER
	OC43	VLLERVYKHL	N---NYDCV	VTTLISAGIF	SVPSDVSLTY	LLGTAKKQV	LVSNNQEDFD
25	BoCoV	ALLERVYKHL	N---KYDCV	VTTLISAGIF	SVPSDVSLTY	LLGTAKKQV	LVSNNQEDFD
	MHV	SFLERAYQHI	N---KCDDV	VTTLISAGIF	SVPTDVSLTY	LIGVVTKNVI	LVSNNQEDFD
	AIPV	NFLILEWRDG	N---CWISS	ATVLLQAAKI	RFGKFLTEAW	AKLLGGDPTD	FVAWCYASCT
	SARS CoV	GTVSWNLREM	LAHAETRKL	MPICMDVRAI	MATIQRYKYG	IKIQEGIVDY	GVRFFFTYSK
30		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		1565	1575	1585	1595	1605	1615
	EMCR	QAVLKFLDGL	DLTPVID---	-----DQDVV	-----	-KPERVEGNF	SFFDCG---V
	229E	CKVKDFVSG	VNVQKVE---	-----QPKIE	PKPVSIVKVA	PKPYRVDGKF	SYFTED---L
35	PEDV	EATIKYMDGL	VDAIFKEALV	DTTPVQEDVQ	QVSQKPVLEN	FEPFRIEGAH	AFYECNPEGL
	TGEV	QTIENFFS---	-----	-----	-----	-----	-----
	OC43	LISKQITAV	EG-----	-----	-----	-----	-----T
	BoCoV	LISKQITAV	EG-----	-----	-----	-----	-----T
	MHV	VIEKQVTSI	AG-----	-----	-----	-----	-----T
40	AIPV	AKVGDFSDAN	-----	-----	-----	-----	-----
	SARS CoV	EPVASIITKL	N-----	-----	-----	-----	-----S
		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		1625	1635	1645	1655	1665	1675
45	EMCR	NALDGD-IYL	LFTNSILMLD	KQGQLLDTKL	NGILQQAVID	YLATVKTVA	GNLVKLVVE-
	229E	LCVADDKPIV	LFTDSMLTLD	DRGLALDNAL	SGVLSAAIKD	CVDINKAIPS	GNLKFDIG-
	PEDV	MSLGAD-KLV	LFTNSNLDFC	SVGKCLNDVT	SGALLEAINV	FKKSNTKTPA	GNCVTLDGAN
	TGEV	-----	-----	-----	-----	-----	-----
	OC43	KKLAARLSFN	VGRSIVYETD	ANKLILIN---	-DVAFVSTFN	VLQDVLSLRH	DIALDDDDART
50	BoCoV	KKLAARLSFN	VGRSIVYETD	ANKLILIN---	-DVAFVSTFN	VLQDVLSLRH	DIALDDDDART
	MHV	KALSLQLAKN	LCRDVKFETN	ACDSLFS---	-DSCFVSSYD	VLQDEVELLRH	DIALDDDDARV
	AIPV	-----	-----	-----	-----	-----	-----
	SARS CoV	LNEPLVTMPI	GYVTHGENLE	EAARCMR---	-SLKAPAVVS	VSSPDAVTTY	NGYLTSSSKT
		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		1685	1695	1705	1715	1725	1735
55	EMCR	SCTIYMCVVP	SI-NDLSFDK	NLGRVCVRKLN	RLKTCVIANV	PAIDVLKKLL	SSLTLTVKFV
	229E	SVVVMCVVP	SE-KDKHLDN	NVQRCTRKLN	RLMCDIVCTI	PADYILPLVL	SSLTCNVSVF
	PEDV	MISITMVVLP	FD-GDANYDK	NYARAVVQVS	KLKGLVLAV	DDATLYSKLS	---HLSVLGFV
	TGEV	-----CSIP	-----	-----	-----	-----	-----
60	OC43	FVQSNVDVVP	EG-WRVVNKF	YQINGVRTVK	YFECTGGIDI	CSQDKVFGYV	QQGIFNKATV
	BoCoV	FVQSNVDVVP	EG-WRVVNKF	YQINGVRPVK	YFECPPGIDI	CSQDKVFGYV	QQGSFNKATV
	MHV	FVQAHMDNLP	AD-WRLVNKF	DSVDGVRTVK	YFECPPGIFV	SSQGGKFGYV	QNGSFKVASV
	AIPV	TNAFLKKRVS	CN-----	-----	-----	-----	-----
65	SARS CoV	SEEHFVETVS	LAGSYRDSY	SGQRTGELVE	FLKRGDKIVY	HTLESPEVEFH	LDG--EVLSL
		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		1745	1755	1765	1775	1785	1795
70	EMCR	VESNMVDVND	CFKNDNVVLK	ITEDGINVKD	VVVESSKSLG	KQLG-VVSDG	VDSFEGVLP-
	229E	GELKAAEA---	-----KVITIK	VTEDGVNVHD	VTVTTDKSFE	QQVG-VIADK	DKDLSGAVPS
	PEDV	STPDDVER---	FYANKSVVVK	VTEDTRSVKA	VKVESTATYG	QQIG-PCLVN	DVVVTDNKP-
	TGEV	-----	-----VN	VTEDNVNHER	VSVSFDKTYG	EQLKGTVVVK	DKDVTNQLPS
	OC43	AQIKALFLD---	-----KVDIL	LTVDGVNFTN	RFVVPVGSFSG	KSLG-NVFC	GVNVTKHKCD
	BoCoV	AQIKALFLD---	-----KVDIL	LTVDGVNFTN	RFVVPVGSFSG	KSLG-NVFC	GVNVTKHKCD
75	MHV	SQIRALLAN---	-----KVDVL	CTVDGVNFTN	CCVAEGEVFG	RTLG-SVFC	GINVTKVRCS
	AIPV	-----	-----IKSYE	LRGLEACIQP	VRATNLLHFK	TQYSNCPCTG	ANNTDEVIEA
	SARS CoV	DKLKSLLSLR	--EVRTIKVF	TTVDNTNLHT	QLVDMSTYTG	QQFG-PYILD	GADVTKIKPH
		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		1805	1815	1825	1835	1845	1855
80	EMCR	-INTDTVLSV	APEVDWVAFY	GFEKAALFAS	LDVKPYG---	-----YPNDF	VGGFRVLGTT

44/87

229E	DLNTSELLTK	AIDVDWVEFY	GFKDAVTFAT	VDHSAFA---	-----YESAV	VNGIRVLKTS
PEDV	-VVADVVAKV	VPNANWDSHY	GFDKAGEFHM	LDHTGFT---	-----FPSEV	VNGRRVIKTT
TGEV	AFDVGQKVIK	AIDIDWQAHY	GFRDAAAFSA	SSHDAK---	-----FEVVV	HSNFIVHKQT
OC43	INYKGVFFQ	FDNLSSSEDLK	AVRSSFNEDQ	KELLAYYNML	VNCFKWQVVV	NGKYFTFKQA
BoCoV	INYKGVFFQ	FDNLSSSEDLK	AVRSSFNEDQ	KELLAYYNML	VNCSKWQVVV	NGKYFTFKQA
MHV	AIHKGKVFQ	YSGLSAADLV	AVTDAFGFDE	PQLLKYYNML	G-MCKWPVVV	CGNYFAFKQS
AIPV	SLPYLLLFAT	DGPATVDCDE	DAVGTVVFFVG	STNSGHICY--	-----TOA	AGQAFDNLAK
SARS CoV	VNHEGKTFV	LPSDDTLRSE	AFEYHTLDE	SFLGRYSAL	NHTKKWKFPQ	VGGLTSIKWA
10	1865	1875	1885	1895	1905	1915
EMCR	DNNCWVNATC	IILQYLKPTF	KSKGLNVLWN	KEVTGDVGPF	VSFIYFITMS	SKGQKGDAAE
229E	DNNCWVNAV	IALQYSKPHF	ISQGLDAAWN	KEVLGDVEIF	VAFVYYVARE	MKGDKGDAED
PEDV	DNNCWVNVTC	LQLQFARFRF	KSAGLOAMWE	SYCTGDVAME	VHWLYWLTGV	DKGQPSDSEN
15	TGEV	DNNCWINAIC	LALQRLKPQW	KFPQVGRGLW	EFLERKTQGF	VHMLYHISGV
OC43	NNNCFVNVSC	LMLQSLHLTF	KIVQWQEAWL	EFRSGRPARF	VALVLAKEGF	KFGDPADSRD
BoCoV	NNNCFVNVSC	LMLQSLHLTF	KIVQWQEAWL	EFRSGRPARF	VSLVLAKEGF	KFGDPADSRD
MHV	NNNCFVNVSC	LMLQSLHLTF	KIVQWQEAWL	EFRSGRPARF	VSLVLAKEGF	KFGDPADSRD
AIPV	DRKFGKSPY	ITAMYTRFAF	KN-ETSLPVA	KQSKGKSKSV	KEDVSNLATS	SKASFDNLTD
20	SARS CoV	DNNCYLSSVL	LALQOLEVKF	NAPALQEAYY	RARAGDAANF	CALILAYSNK
25	1925	1935	1945	1955	1965	1975
EMCR	ALSKLSEYLI	S-----	DSIVTLE	QYSTCDIC--	-----	-----
229E	TLTKLSKYLA	N-----	EAQVQLE	HYSSCVECDA	K-----	-----
PEDV	ALNMLSKYIV	P-----	AGSVTIE	RVTHDGCC--	-----	-----
TGEV	MLHKLGLDMD	N-----	DCEIIVT	HTTACDKC--	-----	-----
OC43	FLRVVFSQVD	LTGAICDF-E	IACKCGVKQE	QRTGLDAVMH	FGTLSREDLE	IGYTVDCSCG
BoCoV	FLRVVFSQVD	LTGAICDF-E	IACKCGVKQE	QRTGLDAVMH	FGTLSREDLE	IGYTVDCSCG
30	MHV	FMVVLVREAD	LSGATCDF-E	FVCKCGVKQE	QRKGVDAVMH	FGTLDKGDLA
AIPV	FEQWYDSNIE	ES-----	LKVOE	SPDNEDKY--	-----	-----
SARS CoV	TMTHLLQHAN	LESARKVLNV	VCKHCGQKTT	TLTGVEAVMY	MGTLSYDNLK	TGVSIPCVCG
35	1985	1995	2005	2015	2025	2035
EMCR	-----	-----	KSTVVEVKS	VVCASVLKDG	-----	CDVGFCPHRH
229E	-----	-----	F KNSVASINSA	IVCASVKRDG	-----	VQVGCYVHGI
PEDV	-----	-----	-CSKRUVTAP	VVNASVLKLG	-----	VEDGLCPHGL
TGEV	-----	-----	-AKVEKVFGP	VVAAPLAING	-----	TDE-TCVHGV
40	OC43	KKLIHCVRFD	VP--FLICSN	TPASVKLPKG	VGSANIFIGD	KVG-HYVHVK
BoCoV	KKLIHCVRFD	VP--FLICSN	TPASVKLPKG	VGSANIFIGD	KVG-HYVHVK	CEQSYQLYDA
MHV	KKLIHCVRFD	VP--FLICSN	TPASVKLPKG	VGSANIFIGD	KVG-HYVHVK	CEQSYQLYDA
AIPV	NKLVHCTQLN	VP--FLICSN	RPEGKPLPDD	VVAANIFTGG	SLG-HYTHVK	CKPKYQLYDA
SARS CoV	RDATQYLVOQ	ESSFVMSAP	PAEYKLQOGT	FLCANETGNN	YQCGHYTHIT	AKETLYRIDG
45	2045	2055	2065	2075	2085	2095
EMCR	KLRSRVKFN	G-----	-----	-----	-----	-RVVITNVEG
229E	KYYSRVRSVR	G-----	-----	-----	-----	-RAIVSVEQ
PEDV	NYIGKVVVVK	G-----	-----	-----	-----	-TTIVVNVGK
TGEV	SVNVKVTQIK	G-----	-----	-----	-----	-TVAITSLIG
OC43	SNVKKVTDVT	GKLSDCLYLK	NLKQTFKSVL	TTYLDDVVK	IEYKPDLSQY	YCDGGKYITQ
BoCoV	SNVKKVTDVT	GKLSDCLYLK	NLKQTFKSVL	TTYLDDVVK	IEYKPDLSQY	YCDGGKYITQ
50	MHV	CNVSKVSEAK	GNFTDCLYLK	NLKQTFSSKL	TTYLDDVVK	VEYNPDLSQY
AIPV	FIYKLTPTD	EN-----	-----	-----	-----	-S-----
SARS CoV	AHLTKMSEYK	GPVTDVIFYE	TSYTTTIKPV	SYKLDGVYTT	EIEPKLDGYY	RKDNAYYTEQ
55	2105	2115	2125	2135	2145	2155
EMCR	PIISQPSKLL	NGIA--YTTF	S-----	-----GSFD	NGHYVVDAA	NNAVYDGARL
229E	LEPCAQSRL	SGVA--YTAF	S-----	-----GPVD	KGHYTVYDTA	KKSMYDGDRE
PEDV	PVVAPSHLFL	KGVS--YTTF	LDN-----	-----GNGV	VGHYTVFDHG	TGMVHDGDAF
TGEV	PIIG---EVL	EATG--YICY	S-----	-----GSNR	NGHYTYDNR	NGLVDAEKA
OC43	RIIKAQFKTF	EKVDGVYTNF	KLIGHTVCDS	LNA-KLGFDS	SKEFVEYKIT	EWPTATGDVV
BoCoV	RIIKAQFKTF	EKVDGVYTNF	KLIGHTVCDS	LNA-KLGFDS	SKEFVEYKIT	EWPTATGDVV
65	MHV	PIIKAQFRFF	EKVEGVYTNF	KLIGHVSIAEK	FNA-KLGFDC	NSPFEYKIT
AIPV	DAISLKAIWV	EGNANFVVGH	PN-----	-----YYSKS	LHIPTFWENA	ENFVKMGDKI
SARS CoV	PIDLVPQPL	PNAS--FDNF	KLTCSENTKFA	DDLQMTGFT	KPASRELSVT	FFPDLDGDDV
70	2165	2175	2185	2195	2205	2215
EMCR	FASD-----	-----	-----	-----	-----LSTLAVTA	IVVVGCCVTS
229E	VKHD-----	-----	-----	-----	-----LSLLSVTS	VVMVGGYVA-
PEDV	VPGD-----	-----	-----	-----	-----LNVSPVTN	VVVSEQTAVV
TGEV	YHFN-----	-----	-----	-----	-----RDLLQVTT	AIASNFVVKK
OC43	LATDDLIVKR	YERGCITFGK	PVIWLS----	-----HEKASL	NSLTYFNRPS	LVDNKKFDVL
BoCoV	LATDDLIVKR	YERGCITFGK	PVIWLS----	-----HEQASL	NSLTYFNRPS	LVDNKKFDVL
75	MHV	LASDDLIVSR	YSGGCVTFFG	PVIWLG----	-----HEEASL	NSLTYFNRPS
AIPV	GGVT-----	-----	-----	-----	-----MGLWRAEH	LNKPNLERIF
80	SARS CoV	AIDYRHYSAS	FKKGAKLLHK	PIVWHINQAT	TKTTFKENTW	CLRCLWSTKP
						VDTNSFEVL



46/87

BoCoV	ALISIQILTT	WPELLMLST	LHWSVRLVS	LANMLPAHVF	MRFYIIASF	IKLFSLEFHV
MHV	GLIGMQLLTT	WPEFFMLET	MHWSAREFVF	VANMLPAFTL	LRFYIVVTAM	YKIFCLCRHV
AIPV	VAGFVLIICYC	VKYLVLNSTV	LQTVGVCLDW	FVQTVFSHEN	FMGAGFYFWL	FYKIYIQVHH
SARS CoV	SAIMQVFFGY	FASHFISNS-	--WLMWFIIS	IVQMAPVSAM	VRMYIFFASF	YYIWKSVYHI
5	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	2645	2655	2665	2675	2685	2695
EMCR	IVGCNNADCV	ACSKSARLKR	VPLQTIINGM	HKSFYVNANG	GTGFCNKHNF	FCVNCDSFGP
229E	LEGCENPDCI	ACSKSARLKR	FPVNTIVNGV	QRSFYVNANG	GSKFCKKHRF	FCVDCDSYGY
10 PEDV	CLGCDKASCV	ACSKSARLKR	VPVQTIQFQT	SKSFYVHANG	GSKFCKKHNH	FCLNCDSYGP
TGEV	VFACSNPSCK	TCSRTARQTR	IPIQVVVNGS	MKTVYVHANG	TGKFCCKHNF	YCKNCDSYGF
OC43	AYGCSKSGCL	FCYKRNRSR	VKSTIVVGM	IRYDVVMANG	GTGFCSKHQW	NCIDCDSYKP
BoCoV	AYGCSKSGCL	FCYKRNRSR	VKSTIVVGM	IRYDVVMANG	GTGFCSKHQW	NCIDCDSYKP
MHV	MYGCSRPGCL	FCYKRNRSR	VKSTIVVGT	LRYDVVMANG	GTGFCAKHQW	NCLNCDSAFGP
15 AIPV	ILYCKDVTCE	VCKRVARSNR	QEVSVVVGGR	KQIVHVYTNS	GYNFCKRHNH	YCRNCDDYGH
SARS CoV	MDGCTSSTCM	MCYKRNRRAT	VECTTIVNGM	KRSFYVYANG	GRGFCKTHNW	NCLNCDTFCT
	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	2705	2715	2725	2735	2745	2755
20 EMCR	GNTFINGDIA	RELGNVVKTA	VQPTAPAYVI	IDKVDVFNFG	YRLYSGDTFW	RYDFDITESK
229E	GSTFITPEVS	RELGNITKTN	VQPTGPAYVM	IDKVEFENG	YRLYSCETFW	RYNFDITESK
PEDV	GCTFINDVIA	TEVGNVVKLN	VQPTGPATIL	IDKVEFSNGF	YRLYSGDTFW	KYNFDITDSK
TGEV	ENTFICDEIV	RDLSNSVKQT	VYATDRSHQE	VTKEVCSGDF	YRFYVGDEFT	SYDYDVKKHK
OC43	GNTFITVEAA	LDLSKELKRP	IQPTDVAYHT	VTDVQVQVCS	MRLFYDRDGO	RTYDDVNASL
25 BoCoV	GNTFITVEAA	LDLSKELKRP	IQPTDVAYHT	VTDVQVQVCS	MRLFYDRDGO	RTYDDVNASL
MHV	GNTFITHEAA	ADLSKELKRP	VNPTDSAYYL	VTEVQVQVCS	MRLFYERDGO	RYVDDVSASL
AIPV	QNTFMSPEVA	GELSEKLKRP	VKPTAYAYHV	VDEACLVDDE	VNLKYKAATP	GKDSASSAVK
SARS CoV	GSTFISDEVA	RDLSLQFKRP	INPTDQSSYI	VDSVAVKNGA	LHLYFDKAGO	KTYERHPLSH
	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	2765	2775	2785	2795	2805	2815
30 EMCR	YSCKEVLKN-	-----	CNVLENFIVY	NNSGS--NIT	QIKNACVYFS	QLLCEPIKLV
229E	YSCKEVFKN-	-----	CNVLDDEFIV	NNNGT--NVT	QVKNASVYFS	QLLCRPIKLV
PEDV	YTCKEALKN-	-----	CSIIITDEFIV	NNNGS--NVN	QVKNACVYFS	QMLCKPVKLV
35 TGEV	YSSQEVVLS-	-----	MILLDDFIVY	SPSGS--ALA	NVRNACVYFS	QLIGKPIKIV
OC43	FVDYSNLLHS	-----KV	KSVPNMHVVV	VENDA--DKA	NFLNAAVFYA	QSLFRPILMV
BoCoV	FVDYSNLLHS	-----KV	KSVPNMHVVV	VENDA--DKA	NFLNAAVFYA	QSLFRPILMV
MHV	FVDMNGLLHS	-----KV	KGVPEHVVV	VENEA--DKA	GFLNAAVFYA	QSLYRPMLLV
AIPV	CFSVTDFLKK	AVFLKEALKC	EQISNDGFIV	CNTQSAHALE	EAKNAAIYYA	QYLCKPILIL
40 SARS CoV	FVNLDNLRAN	-----NT	KGSLPINVIV	FDGKSKCDES	ASKSASVYYS	QLMCQPILLL
	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	2825	2835	2845	2855	2865	2875
45 EMCR	NSELLSTLS-	-VDFNGVLHK	AYVDVLCNSF	FKELTANMSM	AECKATLGLT	-----
229E	DSELLSTLS-	-VDFNGVLHK	AYIDVLRNSF	GKDLNANMSL	AECKRALGLS	-----
PEDV	DSALLASLS-	-VDFGASLHS	AFVSVLNSF	GKDLSSCNM	QDCKSTLGF	D-----
TGEV	NSDLLEDLS-	-VDFKGALFN	AKKNVIKNSF	NVDVSECKNL	DECYRACNLN	-----
OC43	DKNLITTANT	GTSVTETMFD	VYVDTFLSMF	DVDKKSLLNAL	IATAHSSIKQ	GTQIYKVLDT
50 BoCoV	DKLITTANT	GTSVTETMFD	VYVDTFLSMF	DVDKKSLLNAL	IATAHSSIKQ	GTQICKVLDT
MHV	EKKLITTANT	GLSVSQTMEF	LYVDSLLGVL	DVDRKSLTSF	VNAAHNSLKE	GVQLEQVMDT
AIPV	DQALYEQLVV	-EPVSKSVID	KVCSILSSII	SVDTAALNYK	AGTLRDALLS	-----
SARS CoV	DQVLVSDVGD	STEVSVKMF	AYVDTFSATF	SVPMEKLKAL	VATAHSELAK	GVALDGVLST
	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	2885	2895	2905	2915	2925	2935
55 EMCR	-----	-VSDDDF	VSAVANAHRY	DVLLSDLSFN	NFFISYAKPE	DK-LSVYDIA
229E	-----	-ISDHEF	TSATSNARHC	DVLLSDLSFN	NEVSSYAKPE	EK-LSAYDLA
PEDV	-----	-VPLDTF	NAAVABAHRY	DVLLTDMSEN	NFTTSYAKPE	EK-FPVHDLA
TGEV	-----	-VSFSTF	EMAVNNAHRY	GILITDRSFN	NFWPSKVKPG	SSGVSAMDIG
60 OC43	FLSCARKSCS	IDSVDVTKCL	ADSVMSAVSA	GLELTDESCN	NLVPTYLKSD	N--IVAADLG
BoCoV	FLSCARKSCS	IDSVDVTKCL	ADSVMSAVSA	GLELTDESCN	NLVPTYLKSD	N--IVAADLG
MHV	FIGCARRKCA	IDSVDVTKSI	TKSIMSAVNA	GVDFDESCN	NLVPTYVKSD	T--IVAADLG
AIPV	-----	-ITKDEEA	VDMAIFCHNH	DVDYTGDFGT	NVIPSYGIDT	G-KLTPRDRG
65 SARS CoV	FVSAARQG-V	VDTDVDTKDV	IECLKLSHHS	DLEVTDGDCN	NFMLTYNKVE	N--MTPRDLG
	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	2945	2955	2965	2975	2985	2995
70 EMCR	CCMRAGSKVV	NHNVLIKESI	PIVWGVKDFN	TLSQEGKKYL	VKTTKAKGLT	FLLTFNDNOA
229E	CCMRAGAKVV	NANVLTKDQT	PIVWHAKDFN	SLSAEGKKYI	VKTSKAKGLT	FLLTINENQA
PEDV	TCMRVGAKIV	NHNVLVKDSI	PVVWLVRDFI	ALSEETRKYI	IRTTKVKGIT	FLLTFNDCRM
TGEV	KCMSTDAKIV	NAKVLTORGK	SVVWLSQDFA	ALSSTAQKVL	VKTFVEEGVN	FSLTFNAVGS
OC43	VLIQNSAKHV	QGNVAKIAGV	SCIWSVDAFN	QFSSDFQHKL	KKACKCTGLK	LKLTYNKQMA
BoCoV	VLIQNSAKHV	QGNVAKIAGV	SCIWSVDAFN	QFSSDFQHKL	KKACKCTGLK	LKLTYNKQMA
MHV	VLIQNNAKHV	QANVAKAANV	ACIWSVDAFN	QLSADLQHRL	RKACSKTGLK	IKLTYNKQEA
75 AIPV	FLINADASIA	NLRVKN--AP	PVVWKFSELI	KLSDSCLKYL	ISATVKSQVR	FFITKSGAKQ
SARS CoV	ACIDCNARHI	NAQVAKSHNV	SLIWNVKDYM	SLSEQLRKQI	RSAAKKNIP	FRITCATTRQ
	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	3005	3015	3025	3035	3045	3055
80 EMCR	ITQVP----	TSIVAKQGAG	-----	FKRTYNFLWY	VCLFVVALFI	GVSFID----



47/87

229E	VTQIP----	A	TSIVAKQGAG	D-----	AGHSLTWLWL	LCGLVCLIQF	YLCFFMPY--
PEDV	HTTIP----	T	VCIANKKGAG	LP-----	S	FSKVKKFFWF	LCLFIVAAFF
TGEV	DDDLPYERFT		ESVSPKSGSG			FFDVITQLKQ	IVILVVFVIF
OC43	NVSVL----	T	TPFSLKGGAV	FS-----	Y	FVYVCFVLSL	VCFIGLWCLM
BoCoV	NVSVL----	T	TPFSLKGGAV	FS-----	Y	FVYVCFVLSL	VCFIGLWCLM
MHV	NVPIL----	T	TPFSLKGGAV	FS-----	K	VLQWLFVVNL	ICFIVLWALM
AIPV	VIACHT--QK		LLVEKKAGGI	VSGTFKCFKS		YFKWLLIFYI	LFTACCSGYI
SARS CoV	VUNVI----	T	TKISLKGKGI	VS-----	T	CFKMLKATL	LCVLAALVCY
10	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	3065	3075	3085	3095	3105	3115	
EMCR	-YTTTTSFHF	GYDFKYIENG	QLKVFEAPLH	CVRNVFDNFN	QWHEAKFGV	TTNSD-KCPI	
229E	FMVDIVSSFE	GYDFKYIENG	QLKNFEAPLK	CVRNVFENFE	DWHYAKFGFT	PLNKQ-SCPI	
PEDV	-FSTQVSSDS	DYDFKYIESG	QLKTFDNPLS	CVHNVFINFD	QWHDAGFGFT	PVNNP-SCPI	
TGEV	ATQSYIESAE	GYDYMVIKNG	IVQPFDDTIS	CVHNTYKGGF	DWFKAKYGF	PTFGK-SCPI	
OC43	--KSDFOQLPV	YASYKVLONG	VIRDVSVEDV	CFANKFEQFD	QWYESTFGLS	YYSNSMACPI	
BoCoV	--KSDFOQLPV	YASYKVLONG	VIRDVSVEDV	CFANKFEQFD	QWYESTFGLS	YYSNSMACPI	
MHV	--KSDMQPLPL	YASFYKVIDNG	VLRDVTVTDA	CFANKFIOFD	QWYESTFGLV	YYRNSRACPV	
AIPV	PMYDVNSTLH	VEGFKKVIDKG	VLRDVTVTDA	CFANKFIOFD	QWYESTFGLV	YYRNSRACPV	
SARS CoV	--IHDGYTNE	IIGYKAIQDG	VTRDIISTDD	CFANKHAGFD	AWFSGRGGSY	KNDKS--CPV	
25	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	3125	3135	3145	3155	3165	3175	
EMCR	VVG---VSEI	INVVPGVPTN	VYLVG----	-----KTLV	FTLQAAFGNT	GVCYDFDGV	
229E	VVG---VSEI	VNTVAGIPSN	VYLVG----	-----KTLV	FTLQAAFGNT	GVCYDFDGV	
PEDV	VVG---VSEI	VNTVAGIPSN	VYLVG----	-----KTLV	FTLQAAFGNT	GVCYDFDGV	
TGEV	VVG---VSEI	VNTVAGIPSN	VYLVG----	-----KTLV	FTLQAAFGNT	GVCYDFDGV	
OC43	VVG---VSEI	VNTVAGIPSN	VYLVG----	-----KTLV	FTLQAAFGNT	GVCYDFDGV	
BoCoV	VVG---VSEI	VNTVAGIPSN	VYLVG----	-----KTLV	FTLQAAFGNT	GVCYDFDGV	
MHV	VVG---VSEI	VNTVAGIPSN	VYLVG----	-----KTLV	FTLQAAFGNT	GVCYDFDGV	
AIPV	VVG---VSEI	VNTVAGIPSN	VYLVG----	-----KTLV	FTLQAAFGNT	GVCYDFDGV	
SARS CoV	VVG---VSEI	VNTVAGIPSN	VYLVG----	-----KTLV	FTLQAAFGNT	GVCYDFDGV	
30	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	3185	3195	3205	3215	3225	3235	
EMCR	TS-----DK	CIFNSACTRL	EGLGGD-NVY	CYN-TDLIEG	SKPYSILOPN	AYYKYDVKN-	
229E	TP-----EK	CIFNSACTRL	EGLGGD-NVY	CYN-TALMEG	SLPYSSIQAN	AYYKYDNGN-	
PEDV	DK-----GA	CIFNSACTRL	EGLGGD-NVY	CYN-TALMEG	SLPYSSIQAN	AYYKYDNGN-	
TGEV	VSKD-SYFDT	CVFNTACTTL	TGLGGT-IVY	CAK-QGLVEG	AKLYSELAPH	SYKMDVGN-	
OC43	ISYSNFIYASG	CVLSSACTMF	TMDGSPQPY	CYT-EGLMQN	ASLYSSLVPH	VRYNLANAKG	
BoCoV	ISYSNFIYASG	CVLSSACTMF	TMDGSPQPY	CYT-EGLMQN	ASLYSSLVPH	VRYNLANAKG	
MHV	IPYDNFIYASG	CVLSSACTMF	TMDGSPQPY	CYT-EGLMQN	ASLYSSLVPH	VRYNLANAKG	
AIPV	TEG-SFYTSI	ALFSARCLYL	TASNTP-QLY	CFNGDNDARG	ALPFGSIIIP	RVYFQPNVGR	
SARS CoV	IEYSDFATSA	CVLAAECTIF	KDAMGKPEY	CYD-TNLLEG	SISYSELRPD	TRYVLMDS-	
45	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	3245	3255	3265	3275	3285	3295	
EMCR	YVRFPELLAR	GFGRLTIRTL	ATRYCRVGEC	RDSHKGVCFC	FDKWWVNDGR	VD---DGYIC	
229E	FIKLEPIVIAQ	GFGRLTIRTL	ATRYCRVGEC	RDSHKGVCFC	FDKWWVNDGR	VD---DGYIC	
PEDV	AVSLPEIISR	GFGRLTIRTK	AMTYCRVGQC	VQSAEGVCFC	ADRFVYNAE	SG---SDFVC	
TGEV	MVKLPAIR-	GLGLRFVKTQ	ATTYCRVGEC	IDSKAGFCFC	GDWVFDVNE	FG---NGYIC	
OC43	FIRFPEVLRE	GL-VRIVRTR	SMSYCRVGLC	EEADEGICFN	FNGSWVLNND	YYRSLPGTFC	
BoCoV	FIRFPEVLRE	GL-VRIVRTR	SMSYCRVGLC	EEADEGICFN	FNGSWVLNND	YYRSLPGTFC	
MHV	YIRFPEVSE	GI-VRIVRTR	SMTYCRVGLC	EDAEAGVCFC	FNGSWVLNND	YYRSLPGTFC	
AIPV	LIVPQQILHT	PI---VVKFV	SMSYCRVGLC	EDAEAGVCFC	FNGSWVLNND	YYRSLPGTFC	
SARS CoV	IIQFENTYLE	GS-VRVVTF	DAEYCRHGTC	ERSEVGICLS	TSGRWVLNNE	HYRALSGVFC	
50	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	3305	3315	3325	3335	3345	3355	
EMCR	GDGLIDLVLN	VLSIFSSSFS	VVAMSGHMLF	NELFAAFITF	LCFLVTKFKR	VFGDLSVGVF	
229E	GTGLWNLVFN	ILSMFSSSFS	VVAMSGHMLF	NELFAAFITF	LCFLVTKFKR	VFGDLSVGVF	
PEDV	GTGLWNLVFN	ILSMFSSSFS	VVAMSGHMLF	NELFAAFITF	LCFLVTKFKR	VFGDLSVGVF	
TGEV	GNSVLGFFKN	VFKLFNSNMS	VVATSGAMLV	NIIIACLAIA	MCYGVLEKFK	IFGDCTFLIV	
OC43	GRDVEDLIYQ	LFKGLAQPV	FLALTASSIA	GAILAVIVVL	VFYLLIKLKR	AFGDYTSVVF	
BoCoV	GRDVEDLIYQ	LFKGLAQPV	FLALTASSIA	GAILAVIVVL	VFYLLIKLKR	AFGDYTSVVF	
MHV	GRDVEDLIYQ	LFKGLAQPV	FLALTASSIA	GAILAVIVVL	VFYLLIKLKR	AFGDYTSVVF	
AIPV	GRDVEDLIYQ	LFKGLAQPV	FLALTASSIA	GAILAVIVVL	VFYLLIKLKR	AFGDYTSVVF	
SARS CoV	GRDVEDLIYQ	LFKGLAQPV	FLALTASSIA	GAILAVIVVL	VFYLLIKLKR	AFGDYTSVVF	
60	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	3365	3375	3385	3395	3405	3415	
EMCR	TVVCATLINN	ISYVVTQN-L	FEMLLYAILY	FVETRTVR--	YAWIWHIAYI	VAYFLLIPWW	
229E	TVVVAVLLNN	VSIVVTQN-L	VTMIAYAILY	FFATRSRLR--	YAWIWHIAYI	VAYFLLIPWW	
PEDV	TVGACTLLNN	VSIVVTQN-L	VTMIAYAILY	FFATRSRLR--	YAWIWHIAYI	VAYFLLIPWW	
TGEV	MIIVTLVNN	VSIVVTQN-T	LGMLGYATLY	FLCTRGVR--	YAWIWHIAYI	VAYFLLIPWW	
OC43	VNVIVWCVNE	MMLEVFQVYP	FEMIIYAIVY	YFITRKLAA--	YAWIWHIAYI	VAYFLLIPWW	
BoCoV	VNVIVWCVNE	MMLEVFQVYP	FEMIIYAIVY	YFITRKLAA--	YAWIWHIAYI	VAYFLLIPWW	
MHV	VNVIVWCVNE	MMLEVFQVYP	FEMIIYAIVY	YFITRKLAA--	YAWIWHIAYI	VAYFLLIPWW	
AIPV	VNVIVWCVNE	MMLEVFQVYP	FEMIIYAIVY	YFITRKLAA--	YAWIWHIAYI	VAYFLLIPWW	
SARS CoV	VNVIVWCVNE	MMLEVFQVYP	FEMIIYAIVY	YFITRKLAA--	YAWIWHIAYI	VAYFLLIPWW	
70	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	3365	3375	3385	3395	3405	3415	
EMCR	TVVCATLINN	ISYVVTQN-L	FEMLLYAILY	FVETRTVR--	YAWIWHIAYI	VAYFLLIPWW	
229E	TVVVAVLLNN	VSIVVTQN-L	VTMIAYAILY	FFATRSRLR--	YAWIWHIAYI	VAYFLLIPWW	
PEDV	TVGACTLLNN	VSIVVTQN-L	VTMIAYAILY	FFATRSRLR--	YAWIWHIAYI	VAYFLLIPWW	
TGEV	MIIVTLVNN	VSIVVTQN-T	LGMLGYATLY	FLCTRGVR--	YAWIWHIAYI	VAYFLLIPWW	
OC43	VNVIVWCVNE	MMLEVFQVYP	FEMIIYAIVY	YFITRKLAA--	YAWIWHIAYI	VAYFLLIPWW	
BoCoV	VNVIVWCVNE	MMLEVFQVYP	FEMIIYAIVY	YFITRKLAA--	YAWIWHIAYI	VAYFLLIPWW	
MHV	VNVIVWCVNE	MMLEVFQVYP	FEMIIYAIVY	YFITRKLAA--	YAWIWHIAYI	VAYFLLIPWW	
AIPV	VNVIVWCVNE	MMLEVFQVYP	FEMIIYAIVY	YFITRKLAA--	YAWIWHIAYI	VAYFLLIPWW	
SARS CoV	VNVIVWCVNE	MMLEVFQVYP	FEMIIYAIVY	YFITRKLAA--	YAWIWHIAYI	VAYFLLIPWW	
80	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	3365	3375	3385	3395	3405	3415	
EMCR	TVVCATLINN	ISYVVTQN-L	FEMLLYAILY	FVETRTVR--	YAWIWHIAYI	VAYFLLIPWW	
229E	TVVVAVLLNN	VSIVVTQN-L	VTMIAYAILY	FFATRSRLR--	YAWIWHIAYI	VAYFLLIPWW	
PEDV	TVGACTLLNN	VSIVVTQN-L	VTMIAYAILY	FFATRSRLR--	YAWIWHIAYI	VAYFLLIPWW	
TGEV	MIIVTLVNN	VSIVVTQN-T	LGMLGYATLY	FLCTRGVR--	YAWIWHIAYI	VAYFLLIPWW	
OC43	VNVIVWCVNE	MMLEVFQVYP	FEMIIYAIVY	YFITRKLAA--	YAWIWHIAYI	VAYFLLIPWW	
BoCoV	VNVIVWCVNE	MMLEVFQVYP	FEMIIYAIVY	YFITRKLAA--	YAWIWHIAYI	VAYFLLIPWW	
MHV	VNVIVWCVNE	MMLEVFQVYP	FEMIIYAIVY	YFITRKLAA--	YAWIWHIAYI	VAYFLLIPWW	
AIPV	VNVIVWCVNE	MMLEVFQVYP	FEMIIYAIVY	YFITRKLAA--	YAWIWHIAYI	VAYFLLIPWW	
SARS CoV	VNVIVWCVNE	MMLEVFQVYP	FEMIIYAIVY	YFITRKLAA--	YAWIWHIAYI	VAYFLLIPWW	

		3425	3435	3445	3455	3465	3475
5	EMCR	LLTWFSFAAF	LELLPNVFKL	K-----ISTQL	FEGDKFIGTF	ESAAAGTFVL	DMRSYERLIN
	229E	LCAWYFLAML	TGLPLSLLKL	K-----VSTNL	FEGDKFVGTF	ESAAAGTFVI	DMRSYKLAN
	PEDV	VLMVYAFSAI	FEFMPNLFKL	K-----VSTQL	FEGDKFVGTF	ESAAAGTFVL	DMHAYERLAN
	TGEV	VITAYILVFL	YDLSPLSFKL	K-----VSTNL	FEGDKFVGTF	ESAAAGTFVI	DMRSYETIVN
	OC43	FCLLYIAVVV	SN--HAFWVF	S-----YCRKL	GTSVRSDDGT	EEMALTTFFMI	TKDSYCKLKN
	BoCoV	FCLLYISVVV	SN--HAFWVF	S-----YCRQL	GTSVRSDDGT	EEMALTTFFMI	TKDSYCKLKN
10	MHV	FCCIIYAVVV	SN--HALWLF	S-----YCRKL	GTEVRSDDGT	EEMSLTTFFMI	TKESYCKLKN
	AIPV	LACCLYGFII	YMYTLPFLWC	YGTTKNTRKL	YDGNFVGVNY	DIAAKSTFVI	RGSEFVKLTN
	SARS CoV	ITAIYVFCIS	LKHCHWFFNN	Y----LRKRV	MFNGVTFSTF	EAAALCTFLL	NKEMYKLKRS
		3485	3495	3505	3515	3525	3535
15	EMCR	T--ISPEKLL	NYAASYNKYK	YSGSASEAD	YRCACYAHLA	KAMLDYAKDH	N-DMLYSFPT
	229E	S--ISPEKLL	SYAASYNRYK	YSGNANEAD	YRCACYAYLA	KAMLDERSDH	N-DILYTFPT
	PEDV	S--ISTEKLK	QYASTYNNKY	YSGSASEAD	YRLACFAHLA	KAMMDYASNH	N-DTLYTFPT
	TGEV	S--TSIARIK	SYANSFNKYK	YTGSMGEAD	YRMACYAHLG	KALMDYSVNR	T-DMLYTFPT
	OC43	S--LSDVAFN	RYLSLYNKYR	YSGKMDTAA	YREAACSQLA	KAMDTFTNNN	GSDVLYQPPPT
	BoCoV	S--LSDVAFN	RYLSLYNKYR	YSGKMDTAA	YREAACSQLA	KAMDTFTNNN	GSDVLYQPPPT
	MHV	S--VSDVAFN	RYLSLYNKYR	YSGKMDTAA	YREAACSQLA	KAMDTFTNNN	GNDVLYQPPPT
	AIPV	E--I-GDKFE	AYLSAYARLK	YSGTGSEQD	YLOACRAWLA	YALDQYR-NS	GVEIVYTPPR
	SARS CoV	ETLLPLTQYN	RYLALYNKYK	YFSGALDTS	YREAACCHLA	KALNDFS-NS	GADVLYQPPQ
		3545	3555	3565	3575	3585	3595
30	EMCR	ISYN-STLQS	GLKKMAQPSG	CVERCVVRVC	YGSTVLNGVW	LGDTVTCPRH	VIAPS-TTVL
	229E	VSYG-STLQA	GLRKMAQPSG	FVEKCVVRVC	YGNTVLNGLW	LGDIVMCPRH	VIASN-TTSA
	PEDV	VSYN-STLQA	GLRKMAQPSG	VVEKCIVRVC	YGNMALNGLW	LGDIVMCPRH	VIASS-TTST
	TGEV	VSVN-STLQS	GLRKMAQPSG	LVEPCIVRVS	YGNVNLNGLW	LGDEVICPRH	VIASD-TTRV
	OC43	ASVSTSFLOS	GIVKMVNPTS	KVEPCVVSST	YGNMTLNGLW	LDDKVYCPRH	VICSASDMTN
	BoCoV	ASVSTSFLOS	GIVKMVNPTS	KVEPCIVSST	YGNMTLNGLW	LDDKVYCPRH	VICSASDMTN
	MHV	ASVTTSFLOS	GIVKMVFPTS	KVEPCVVSST	YGNMTLNGLW	LDDKVYCPRH	VICSSADMTD
35	AIPV	YSIGVSRLOS	GFKKLVSPTS	AVEKCIVSST	YRGNLNLGLW	LGDTVYCPRH	VLG---KFSG
	SARS CoV	TSITSVAVLOS	GFRKMAFPST	KVEGCMVQVT	CGTTTLNGLW	LDDTVYCPRH	VICTAEDMLN
		3605	3615	3625	3635	3645	3655
40	EMCR	IDYDHAYSTM	RLHNFVSVSH	G-VFLGVVGV	TMHGSVLRK	VSQSNVHTPK	HVFETLKPGA
	229E	IDYDHEYSIM	RLHNFISIISG	T-AFLGVVGA	TMHGVTLKIK	VSQTNMHTPK	HSFRTLKSGE
	PEDV	IDYDYALSVM	RLHNFISISSG	N-VFLGVVSA	TMRGALLQIK	VQNQNVHTPK	YTYRTVRPGE
	TGEV	INYNEMSSV	RLHNFVSVK	N-VFLGVVSA	RYKGVNLVLK	VNQVNPTPE	HKFKSIKAGE
	OC43	PDYTNLLCRV	TSSDFTVLED	R-LSLTVMYS	QMRGCMVLVT	VTLQNSRTPK	YTFGVVKPGE
45	BoCoV	PDYTNLLCRV	TSSDFTVLED	R-LSLTVMYS	QMRGCMVLVT	VTLQNSRTPK	YTFGVVKPGE
	MHV	PDYSNLLCRV	ISSDFCVMSG	R-MSLTVMYS	QMQGSLVLVT	VTLQNPNTPK	YSEGVVKPGE
	AIPV	DQWNDVNLMA	NNHEFEVTTQ	HGVTLNVVSR	RLKGAVLILQ	TAVANAETPK	YKFIKANCSD
	SARS CoV	PNYEDLLIRK	SNHSFVLQAG	N-VQLRVIGH	SMQNCLLRLK	VDTSNPRTPK	YKFVIRIQPGQ
		3665	3675	3685	3695	3705	3715
50	EMCR	SFNILACYEG	IASGVFGVNL	RNETIKGSF	INGACGSPGY	NVRNDGTVEF	CYLHQIELGS
	229E	GFNILACYDG	CAQGVFGVNM	RTNWTIRGSF	INGACGSPGY	NLRN-GEVEF	VYMHQIELGS
	PEDV	SFNILACYDG	AAAGVYGVNM	RSNYTIRGSF	INGACGSPGY	NINN-GTVEF	CYLHQIELGS
	TGEV	SFNILACYEG	CPGSVYGVNM	RSQGTIKGSF	IAGTCGSGVY	VLEN-GILYF	VYMHQIELGN
	OC43	TFTVLAAYNG	KPQGAHFVMT	RSSYTIKGSF	LCGSCGSGVY	VIMG-DCVKE	VYMHQIELST
	BoCoV	TFTVLAAYNG	KPQGAHFVMT	RSSYTIKGSF	LCGSCGSGVY	VIMG-DCVKE	VYMHQIELST
	MHV	TFTVLAAYNG	KSQGAHFVMT	RSSYTIKGSF	LCGSCGSGVY	VLTG-DSVRE	VYMHQIELST
60	AIPV	SFTIACAYGG	TVVGLYFVMT	RSGNTIRASF	LACGSCSGVF	NIEK-GVUNF	FYMHHLPLN
	SARS CoV	TFSVLACYNG	SPSGVYQCAM	RPNHTIKGSF	LNGSCGSGVF	NIDY-DCVSF	CYMHMELPT
		3725	3735	3745	3755	3765	3775
65	EMCR	GAHVGSDFTG	SVYGNFDDQP	SLQVESANLM	LSDNVVAFLY	AALLNGCR--	----WWLRST



49/87

BoCoV	KCSVEDFNWV	ALSNGFSQVK	SD--LVIDAL	ASMTGVSLET	LLAAIKRLKN	-GFQGRQIMG
MHV	SCSLEEFNVW	AMTNGFSSIK	AD--LVIDAL	ASMTGVTVEQ	LLAAIKRLYS	-GFQGRQILG
AIPV	TVSVDDYNKW	AGDNGFTPFS	TS--TATTKL	SAITGVDVCK	LLRTIMVKNS	-QWGGDPILG
SARS CoV	TTTLNDFNLV	AMKYNYEPLT	QDHVDILGPL	SAQTGIAVLD	MCAALKEILQ	NGMNGRTILG
5	EMCR	YSSLCDEFTL	AEVVKQMYGV	NLQSGK---V	IFGLKTMFLF	SVFFTMFWAE
10	229E	YSSLNDEFSI	NEVVKQMFV	NLQSGK---T	TSMFKSISLF	AGFFVMFWAE
PEDV	HTSLTDEFTT	GEVVRQMYGV	NLQGGY---V	SRACRNVLLV	GSFLTFFWSE	LVSYTKFFWV
TGEV	YGSCLDEFTT	TEVIRQMYGV	NLQAGK---V	KSFFYPIMTA	MTILFAFWLE	FFMYTPFTWI
OC43	SCSFEDELTP	SDVYQQLAGI	KLQSKRTRLF	KGTVCWIMAS	TFLFSCIITA	FVKWTMFMVY
BoCoV	SCSFEDELTP	SDVYQQLAGI	KLQSKRTRLV	KGIVCWIMAS	TFLFSCIITA	FVKWTMFMVY
MHV	SCVLEDELTP	SDVYQQLAGV	KLQSKRTRVV	KGTCCWILAS	TLLFCSIISA	FVKWTMFMVY
15	AIPV	QYNFEDELTP	ESVFNQIGGV	RLQSSFVR--	K--ATSWFWS	RCVLACFLFV
SARS CoV	STILEDEFTT	FDVVRQCSGV	TFQGFKKIV	KGTHHWMLLT	FLTSLILVQ	STQWSLFFV
20	EMCR	NPVILTPIFC	LLLFLSLVLT	MFLKHKFLFL	QVFLLPVIA	TALYNC-VLD
229E	NPGFLTPFMI	LLVALSLCLT	FVVKHKVLEL	QVFLLPISIV	AAIQNC-AWD	YHVTKVLAEK
PEDV	NPGYVTPMFA	CLSLSSLLM	FTLKHKTLEF	QVFLIPALIV	TSCINL-AED	VEVYNLAEH
TGEV	NPTFVSIOLA	VTTLISTVFV	SGIKHKMLFF	MSFVLPVIL	VTAHNL-FWD	FSYESLQSI
OC43	TTNMFISITFC	ALCVIS-LAM	LLVKHKHLYL	TMVITP-VLF	TLLYNN-YLV	VYKHTFRGYV
BoCoV	TTNMLSITFC	ALCVIS-LAM	LLVKHKHLYL	TMVITP-VLF	TLLYNN-YLV	VYKHTFRGYV
MHV	TTHMLGVTLT	ALCFVS-FAM	LLVKHKHLYL	TMFIMP-VLC	TLEFTN-YLV	VYKQSFRLA
25	AIPV	PLKPYVYAAV	ILLMAVLFIS	FTVKHVMAYM	DTFLLPITIT	VIIGVCAEVP
SARS CoV	YENAFLPFTL	GIMAIACAM	LLVKHKHAFI	CLFLLPSLAT	VAYFN---MV	YMPASVWMRI
30	EMCR	FN-YNVSVLQ	MDVQGLVNVL	VCLFVVFELH	---TWRFSEK	ETHWFTYVCS
229E	FD-YNVSVMQ	MDIQGFVNIF	ICLFVALLH-	---TWRFSEK	CTHWCTYLES	LIAYATYFY
PEDV	FD-YHVSIMG	ENAOGLVNIF	VCFVVTILHG	TYTWRFEN-T	PASSVTYVVA	LLTAANYFY
35	TGEV	VENTNTMFLP	VDMQGVMLTV	FCFIVFVYS	VREFTCKQSW	FSLAVTITLV
OC43	YAWLSYVPS	VEYTYTDEVI	YGMILLVGMV	FVTLRSINHD	LFSFIMFVGR	LISVFSWLWK
BoCoV	YAWLSYVPS	VEYTYTDEVI	YGMILLVGMV	FVTLRSINHD	LFSFIMFVGR	VISVSVLWYM
MHV	YAWLSHFVPA	VDYTYMDEVL	YGVVLVAMV	FVTMRINHD	VFSVMFLVGR	LVSLSVMWYF
40	AIPV	VIFLSQWYDP	VVFDTMVPWM	FLPLVLVYAF	KCVQGCYMN	FNTSLLMLYQ
SARS CoV	MTWLELADTS	LSGYRLKDCV	MYASALVLLI	LMTARTVYDD	AARRVWTLMN	VITLVYKVY
45	EMCR	SGD-----	-----FLSL	LVMFLCAISS	DWYIGAIYFR	LSRLIIFSP
229E	SYD-----	-----YVSL	LVMFLCAISS	EWYIGAIYFR	ICRFVGAFLP	V-----SVFSVF
PEDV	ASD-----	-----ILSC	AMTLFASVTC	NWFVGAICYK	VAVYMARLRF	-----TFVALF
TGEV	SDEPWENQI	-----AFCFVNM	LTMIVSLTTK	DWMVVIASR	IAYYIVVCVM	P-S-AFVSDF
OC43	GSN-----	-----LEEEI	LLMLASLFGT	YTWTTLVSLMA	VAKVIAKWVA	VNV-LYFTDI
BoCoV	GSN-----	-----LEEEI	LLMLASLFGT	YTWTTLVSLMA	VAKVIAKWVA	VNV-LYFTDI
MHV	GAN-----	-----LEEEV	LLFLTSLSFGT	YTWTTLVSLMA	TAKVIAKWLA	VNV-LYFTDV
50	AIPV	SSNTLTAYTE	GNWELFFELV	HTTTLVANVS	NSLIGLEVFK	CAKWMLYCN
SARS CoV	GNALD-----	-----QAISM	WALVISVTSN	YSGVVTIMF	LARAIVFCV	EYYPPLFITG
55	EMCR	GDVKLTLLVY	LICGYLVCTY	WGILYWFNRF	FKCTMGVYDF	KVSAAEFKYM
229E	DGVKTVLLFY	MLLGFVSCMY	YGLLYWINRF	CKCTLGVYDF	CVSPAEFKYM	VANGLHAPY
PEDV	GDIKSVMFY	LVLGYETCCF	YGILYWFNRF	EKVSQGVYDY	TVSAAEFKYM	VANGLNAPNG
TGEV	GFMKISIVY	MACGYLFCCY	YGILYWFNRF	TCMTCGVYQF	TVSAAELKYM	TANNLSAPKN
60	OC43	PQIKIVLLCY	LFIFYIISCY	WGLFSLMNSL	FRMPLGVYNY	KISVQELRYM
BoCoV	PQIKIVLLCY	LFIFYIISCY	WGLFSLMNSL	FRMPLGVYNY	KISVQELRYM	NANGLRPPKN
MHV	PQIKIVLLCY	LFIFYIISCY	WGLFSLMNSL	FRMPLGVYNY	KISVQELRYM	NANGLRPPKN
65	AIPV	PQIKIVLLCY	LFIFYIISCY	WGLFSLMNSL	FRMPLGVYNY	KISVQELRYM
SARS CoV	NTLQCMILVY	CFLGYCCTCY	FGLYVWVNVK	FGLTLGKYNE	KVSVDQYRYM	CLHKINPPKT
70	EMCR	PFDAWLSFK	LLGIGGDRCI	KISTVQSKLT	DLKCTNVVLL	GCLSSMNIAA
229E	PFDAWLSFK	LLGIGGDRCI	KISTVQSKLT	DLKCTNVVLL	GCLSSMNIAA	NSSEWAYCVD
PEDV	TLDSLILLSAK	LIGIGGDRNI	KISTVQSKLT	DLKCTNVVLL	GCLSSMNIAA	NSKEWAYCVD
TGEV	AYDAMILSAK	LIGIGGDRNI	KISTVQSKLT	DLKCTNVVLL	GCLSSMNIAA	NSKEWAYCVD
OC43	SFEALMLNFK	LLGIGGVPII	EVSQFQSKLT	DVKCANVLL	NCLQHLHVAS	NSKLWHYCST
BoCoV	SFEALMLNFK	LLGIGGVPII	EVSQFQSKLT	DVKCANVLL	NCLQHLHVAS	NSKLWHYCST
MHV	SFEALMLNFK	LLGIGGVPII	EVSQFQSKLT	DVKCANVLL	NCLQHLHVAS	NSKLWHYCST
75	AIPV	SFEALMLNFK	LLGIGGVPII	EVSQFQSKLT	DVKCANVLL	NCLQHLHVAS
SARS CoV	SIDAFKLNK	LLGIGGKPCI	KVATVQSKMS	DVKCTSVVLL	SVLQQLRVES	SSKLWACQVQ
80	EMCR	LHNKINLCDD	PEKAQGMLLA	LLAFFLSKHS	DFG-----L	DGLIDSYFDN
						SSTLQSVASS

50/87

229E	MHNKINLCDD	PETAQELLLA	LLAFFLSKHS	DFG-----L	GDLVDSYFEN	DSILQSVASS
PEDV	LHNKINLCND	PEKAQEMLLA	LLAFFLSKNS	AFG-----L	DDLLESYFND	NSMLQSVAST
TGEV	LHNEINLCDD	PEIVLEKLLA	LIAFFLSKHN	TCD-----L	SELIESYFEN	TTILQSVASA
OC43	LHNEILATSD	LSVAFELKLAQ	LLIVLFANPA	AVDSKCLTSI	EEVCCDYAKD	NTVLQALQSE
5	BoCoV	LHNEILATSD	LGVAFEKLAQ	LLIVLFANPA	AVDSKCLTSI	EEVCCDYAKD
MHV	LHNEILATSD	LSVAFDKLAQ	LLVVLFFANPA	AVDSKCLASI	EEVSDDYVRD	STVLQALQSE
AIPV	LHNKILASDD	VGECMDNLLG	MLITLFCIDS	TID-----L	SEYCCDILKR	STVLQSVTQE
SARS CoV	LHNDILLAKD	TTEAFEKMVS	LLSVLLSMQG	AVD-----I	NRLCEEMLDN	RATLQAIASE
10		.... ....	.... ....	.... ....	.... ....	.... ....
		4265	4275	4285	4295	4305
EMCR	FVSMPSYIAY	ENARQAYEDA	IANGSS----	SOLIKQLKRA	MNIAKSEFDH	EISVQKKINR
229E	FVGMPSFVAY	ETARQYENYA	VANGSS----	PQIIKQLKKA	MNVAKAEFDR	ESSVQKKINR
PEDV	YVGLPSYVIY	ENARQYEDA	VNNGSP----	PQLVKQLRHA	MNVAKSEFDR	EASTQRKLD
15	TGEV	YAALPSWIAL	EKARADLEEA	KKNDVS----	PQILKQLTKA	FNIAKSDFER
OC43	FVNMAFVEY	EVAKKNLDEA	RFGSGSAN----	QQQLKQLEKA	CNIAKSAYER	DRAVAKKLER
BoCoV	FVNMAFVEY	EVAKKNLDEA	CSSGSAN----	QQQLKQLEKA	CNIAKSAYER	DRAVAKKLER
MHV	FVNMAFVEY	ELAKKNLDEA	KASGSAN----	QQQIKQLEKA	CNIAKSAYER	DRAVAKKLER
AIPV	FSHIPSYAEY	ERAKNLYEKV	LVDSKNGGVT	QQELAYRKA	ANIAKSVFDR	DLAVQKKLDS
20	SARS CoV	FSSLP SYAAY	ATAQEAYEQA	VANGDS----	EVVLKKLKKS	LNVAKSEFDR
		.... ....	.... ....	.... ....	.... ....	.... ....
		4325	4335	4345	4355	4365
EMCR	MAEQAAATQMY	KEARSVNRRKS	KVISAMHSL	FGMLRRRLDMS	SVETVLNLAR	DGVVPLSVIP
229E	MAEQAAAAMY	KEARAVNRKS	KVVSAMHSL	FGMLRRRLDMS	SVDTILNMR	NGVVPLSVIP
PEDV	MAEQAAAQMY	KEARAVNRKS	KVVSAMHSL	FGMLRRRLDMS	SVDTILNLAK	DGVVPLSVIP
TGEV	MAEQAAAAMY	KEARAVNRKS	KVVSAMHSL	FGMLRRRLDMS	SVNTIIDQAR	NGVPLSVIP
OC43	MADLALTNMY	KEARINDKKS	KVVSALQTML	FSMVRKLDNQ	ALNSILDNAV	KGCVPPLNAIP
BoCoV	MADLALTNMY	KEARINDKKS	KVVSALQTML	FSMVRKLDNQ	ALNSILDNAV	KGCVPPLNAIP
30	MHV	MADLALTNMY	KEARINDKKS	KVVSALQTML	FSMVRKLDNQ	ALNSILDNAV
AIPV	MAERAMTMY	KEARVTDRA	KLVSSSLHALL	FSMLKKIDSE	KLVNLFDOAS	SGVVPLATVP
SARS CoV	MADQAMTQMY	KQARSEDKRA	KVTSAMQTML	FTMLRKLDND	ALNNIINNAR	DGCVPPLNAIP
		.... ....	.... ....	.... ....	.... ....	.... ....
		4385	4395	4405	4415	4425
EMCR	ATSASKLTIV	SPDLESYSKI	VCDGSVHYAG	VVWTLNDVKD	NDGRFVHVKE	ITR-----EN
229E	ATSARLVVV	VPDHDSEFKM	MVDGFVHYAG	VVWTLQEVKD	NDGKNVHLKD	VTK-----EN
PEDV	AVSATKLNIV	TSDDISYNRI	QREGCVHYAG	TIWNIIDIKD	NDGKVHVHKE	VTA-----QN
TGEV	AASATRLVVI	TPSLEVFESKI	ROENNVHYAG	AIWTIVEVKD	ANGSHVHLKE	VTA-----AN
40	OC43	SLAANTLNII	VPDKSVYDQV	VDNVVYTYAG	NVWQIQTIQD	SDGTNKQLNE
BoCoV	SLAANTLTII	VPDKSVYDQV	VDNVVYTYAG	NVWQIQTIQD	SDGTNKQLHE	IS-----
MHV	SLTSNTLTII	VPDKQVFDQV	VDNVVYTYAG	NVWHIQSIQD	ADGAVKQLNE	ID-----
AIPV	IVCSNKLTIV	IPDPETWVKC	VEGVHVTYST	VVWNIDTVID	ADGTELHPHS	TSGSLTYCIS
45	SARS CoV	LTAAKLMVV	VPDYGTYNK	CDGNTFTYAS	ALWEIQQVVD	ADSKIVQLSE
		.... ....	.... ....	.... ....	.... ....	.... ....
		4445	4455	4465	4475	4485
EMCR	VEILTWPILI	NCER-----	VVKLQNNNEIM	PGKLRQKPMK	AEG--DGGVL	GDGNALYNTE
229E	QEILVWPLIL	TCER-----	VVKLQNNNEIM	PGKMKVKATK	GEG--DGGIT	SEGNALYNNE
PEDV	AESLSWPLVL	GCER-----	IVKLQNNNEII	PGKLRQRSIK	AEG--DG-IV	GEGKALYNNE
TGEV	ELNLTWPLSI	TCER-----	TTKLQNNNEIM	PGKLRQRAVR	ASATLDGEAF	GSGKALMASE
OC43	-DDCNWPLVI	IANRY-NEVS	ATVLQNNNELM	PAKLKIQVNV	SGP--DQTCN	TPTQCYNNNS
BoCoV	-DDCNWPLVI	IANRH-NEVS	ATVLQNNNELM	PAKLKIQVNV	SGP--DQTCN	TPTQCYNNNS
50	MHV	-VNITWPLVI	AANRH-NEVS	SVVLQNNNELM	PQKLRTQVNV	SGS--DMNCN
AIPV	GANIAWPLKV	NLTRNGHNVK	DDVLQNNNELM	PHGVKTKACV	AGVD-QAHCS	VESKCYNTNI
55	SARS CoV	SPNLAWPLIV	TALRA-N--S	AVKLQNNNELS	PVALRQMSCA	AGTTQTACTD
		.... ....	.... ....	.... ....	.... ....	.... ....
		4505	4515	4525	4535	4545
EMCR	GGKTFMYAYI	SNKADLFVK	WEY-EGG-CN	TIELDSPCRF	MVETPNGPQV	KYLYFVKNLN
229E	GGRAFMYAYV	TTKPGMKYVK	WEH-DSG-VV	TVELEPPCRF	VIDTPTGPQI	KYLYFVKNLN
PEDV	GGRTFMYAFI	SDKPDLRVVK	WEF-DGG-CN	TIELEPPKRF	LVDSPPNGAQI	KYLYFVKNLN
TGEV	SGKSFMAYFI	ASDNNLKYYK	WES-NND-II	PIELEAPLRF	YVDGANGPEV	KYLYFVKNLN
60	OC43	NNGKIVYAIL	SDVDGLKYTK	ILKDDGN-FV	VLELDPPCKF	TVQDAKGLKI
BoCoV	YNGKIVYAIL	SDVDGLKYTK	ILKDDGN-FV	VLELDPPCKF	TVQDVKGGLKI	KYLYFVKGCN
65	MHV	GMGKIVYAIL	SDCDGLKYTK	IVKEDGN-CV	VLELDPPCKF	SVQDVKGGLKI
AIPV	SGNSVVAAIT	SSNPNLKVAS	FLNEAGN-QI	YVDLDPPCKF	GMKVGVKVEV	VYLYFIKNTN
70	SARS CoV	KGGRFVLALL	SDHQDLKWAR	FPKSDGTGTI	YTELEPPCRF	VTDTPKGPVK
		.... ....	.... ....	.... ....	.... ....	.... ....
		4565	4575	4585	4595	4605
EMCR	TLRRGAVLGF	IGATIRLQAG	-KQTELVANS	GLLTACAFSV	DPATTYLEAV	KHGAKFVSNC
229E	NLRRGAVLGY	IGATVRLQAG	-KQTEFVSNS	HLLTHCSFAV	DPAAAYLDAV	KQGAKPVGNC
PEDV	TLRRGAVLGY	IGATVRLQAG	-KQTEQAIN	SLLTLCAFAV	DPAKTYIDAV	KSGHKPVGNC
TGEV	TLRRGAVLGY	IGATVRLQAG	-KPTHEPSNS	SLLTLCAFSV	DPAKAYVDV	KRGMQFVNNC
75	OC43	TLARGWVVG	ISSTVRLQAG	-TATEYASNS	SILSLCAFSV	DPKKTLYLDFI
BoCoV	TLARGWVVG	ISSTVRLQAG	-TATEYASNS	SILSLCAFSV	DPKKTLYLDFI	QGGGTFPIANC
MHV	TLARGWVVG	LSSTVRLQAG	-TATEYASNS	AIRSLCAFSV	DPKKTLYLDYI	QGGGAPVTNC
AIPV	SIVRGMVLGA	ISNVVVLQSK	GHEFTEVDV	GILSLCSFAV	DPADTYCKYV	AAGNQPLGNC
80	SARS CoV	NLNRGMVLGS	LAATVRLQAG	-NATEVPANS	TVLSFCAFAV	DPAKAYKDYL
		.... ....	.... ....	.... ....	.... ....	.... ....
		4565	4575	4585	4595	4605
EMCR	TLRRGAVLGF	IGATIRLQAG	-KQTELVANS	GLLTACAFSV	DPATTYLEAV	KHGAKFVSNC
229E	NLRRGAVLGY	IGATVRLQAG	-KQTEFVSNS	HLLTHCSFAV	DPAAAYLDAV	KQGAKPVGNC
PEDV	TLRRGAVLGY	IGATVRLQAG	-KQTEQAIN	SLLTLCAFAV	DPAKTYIDAV	KSGHKPVGNC
TGEV	TLRRGAVLGY	IGATVRLQAG	-KPTHEPSNS	SLLTLCAFSV	DPAKAYVDV	KRGMQFVNNC
OC43	TLARGWVVG	ISSTVRLQAG	-TATEYASNS	SILSLCAFSV	DPKKTLYLDFI	QGGGTFPIANC
BoCoV	TLARGWVVG	ISSTVRLQAG	-TATEYASNS	SILSLCAFSV	DPKKTLYLDFI	QGGGTFPIANC
MHV	TLARGWVVG	LSSTVRLQAG	-TATEYASNS	AIRSLCAFSV	DPKKTLYLDYI	QGGGAPVTNC
AIPV	SIVRGMVLGA	ISNVVVLQSK	GHEFTEVDV	GILSLCSFAV	DPADTYCKYV	AAGNQPLGNC
80	SARS CoV	NLNRGMVLGS	LAATVRLQAG	-NATEVPANS	TVLSFCAFAV	DPAKAYKDYL

		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		4625	4635	4645	4655	4665	4675
5	EMCR	IKMLSNAGN	GQAITTSVDA	NTNQDSYGGA	SICLYCRAHV	PHP-----SMD	GYCKFKGKCV
	229E	VKMLTNGSGS	GQAITCTIDS	NTTQDTYGGG	SVCICYCRAHV	AHP-----TMD	GFCQYKQKVV
	PEDV	VKMLANGSGN	GQAVTNGVEA	STNQDSYGGA	SVCICYCRAHV	EHP-----SMD	GFCRLKGKVV
	TGEV	VKMLSNAGN	GMAVTNGVEA	NTQQDSYGGA	SVCICYCRCHV	EHP-----AID	GLCRYKQKVV
	OC43	VKMLCDHAGT	GMAITVKPDA	TTSQDSYGGA	SVCICYCRARV	EHP-----DVD	GLCKLRGKVV
10	BoCoV	VKMLCDHAGT	GMAITVKPDA	TTSQDSYGGA	SVCICYCRARV	EHP-----DVD	GLCKLRGKVV
	MHV	VKMLCDHAGT	GMAITIKPEA	TTNQDSYGGA	SVCICYCRSRV	EHP-----DVD	GLCKLRGKVV
	AIPV	VKMLTVHNGS	GFAITSKPSP	TPDQDSYGGA	SVCICYCRAHI	AHPGSVGNLD	GLCQFKGSFV
	SARS CoV	VKMLCTHTGT	GQAITVTPEA	NMDQESFPGA	SCCLYCRCHI	DHP-----NPK	GFCDLKGKVV
15		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		4685	4695	4705	4715	4725	4735
	EMCR	QVPIGCL-DP	IRFCLENNVC	NVCGCWLGHG	CACDRTTIQS	-----	-VDISYLNQEQ
	229E	QVPIGTN-DP	IRFCLNTVVC	KVCGCWLNHG	CTCDRTAIQS	-----	-FDNSYLNES
	PEDV	QVPLGTV-DP	IRFVLENDVC	KVCGCWLNSG	CTCDRSIMQS	-----	-T-----
20	TGEV	QIPTGTQ-DP	IRFCIENEVC	VVCGCWLNNG	CMCDRTSMQS	F-----	TVDQSYLNQEQ
	OC43	QVPVGIK-DP	VSYVLTHDVC	RVCGFWRDGS	CSCVSTDTTV	Q-----	SKDT-----
	BoCoV	QVPVGIK-DP	VSYVLTHDVC	QVCGFWRDGS	CSCVSTDTTV	Q-----	SKDTNFLNGF
	MHV	QVPLGIK-DP	VSYVLTHDVC	QVCGFWRDGS	CSCVGTGSQF	Q-----	SKDTNFLNGF
	AIPV	QIPTTEK-DP	VGFCRLRNVC	TVCQCWIGYG	CQCDSLRQPK	SSVQSVAGAS	DFDKNYLNGY
25	SARS CoV	QIPTTCANDP	VGFTLRNTVC	TVCGMWKGYG	CSCDQLREPL	M-----	QSADASTFLN
30		.... ..	4745				
	EMCR	GVLVQLD					
	229E	GALVPLD					
	PEDV	-----					
	TGEV	GVLVQLD					
	OC43	-----					
	BoCoV	GVRV---					
	MHV	GVQV---					
35	AIPV	GVAURLG					
	SARS CoV	GFAV---					

## c. Putative orf 1b

45		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		5	15	25	35	45	55
	EMCR	-----	-----	RARGSSAARL	EPCN-GTDID	KCVRAFDIYN	KNVSFLGKCL
	229E	-----	-----	-----	EPCN-GTDID	YCVRAFDVYN	KDASFIGKNL
50	PEDV	-----	-----	YGLFK	RVRGSSAARL	EPCN-GTDTQ	HVYRAFDIYN
	TGEV	-----	-----	-----	-----	EPCN-GTDPD	HVSRAFDIYN
	BoCoV	-----	-----	FFKR	VRGTSVDARL	VPCASGLSTD	VQLRAFDICN
	OC43	-----	-----	FFKR	VRGTSVDARL	VPCASGLSTD	VQLRAFDIYN
	MHV	LFLCRHRLPV	SVKRHELFRK	VRGTSVNARL	VPCASGLSTD	VQLRAFDICN	ANRAGIGLYY
55	AIPV	-----	-----	-----	-----	-----	-----
	SARS CoV	-----	-----	-----	TPCGTGTSTD	VVYRAFDIYN	EKVAGFAKFL
60		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		65	75	85	95	105	115
	EMCR	KMNCVREFKNA	-----DL	KDGYFVIKRC	TKSVMHEEQS	MYNLLNFSGA	LAEHDFFTWK
	229E	KSNCVREFKNV	-----DK	DDAFYIVKRC	IKSVMDEHQS	MYNLLKGCNA	VAKHDFFTWH
	PEDV	KVNCVRLKNL	-----DK	HDAFYIVKRC	TKSVMHEEQS	LYSRLEKCGA	IAEHDFFTWK
	TGEV	KTNCSEFRNL	-----DK	HDAFYIVKRC	TKTVMDEHQS	CYNDLKDSGA	VAEHDFFTYK
	BoCoV	KVNCCRFQRV	-----DK	LDQFFVVKRT	DLTIYNREME	CYERVKDCKF	VAEHDFFTFD
65	OC43	KVNCCRFQRV	-----DK	LDQFFVVKRT	DLTIYNREME	CYERVKDCKF	VAEHDFFTFD
	MHV	KVNCCRFQRA	-----DK	LDQFFVVKRT	DLTIYNREME	CYERVKDCKF	VAEHDFFTFD
	AIPV	KRNCARFQEL	RDTEGDNLEY	LDSYFVVKQT	TPSNYEHEKS	CYEDLKS-EV	TADHDFEVEN
	SARS CoV	KTNCCRFQEK	-----NL	LDSYFVVKRH	TMSNYQHEET	IYNLVKDCPA	VAVHDFEFKFR
70		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		125	135	145	155	165	175
	EMCR	DGRVIYGNVS	RHNLTKYTMM	DLVYAMRNFD	EQNCDELKEV	LVLTGCCDNS	-----YFDSKG
	229E	EGRTIYGNVS	RQDLTKYTMM	DLVYAMRNFD	EKDCVFEKEI	LVLTGCCSTD	-----YFEMKN
	PEDV	DGRAIYGNVC	RKDLTEYTMM	DLVYAMRNFD	ENNCDELKSI	LIKVGACEES	-----YFNNKV
	TGEV	EGRCFEGNVA	RRNLTKYTMM	DLVYAMRNFD	EKNCEVLKEI	LVTVGACTEE	-----YFENKD
75	BoCoV	VEGSRVPHIV	RKDLTKYTMM	DLVYALRHFD	RNDCLMLCDI	LSIYAGCEQS	-----YFTKKD
	OC43	VEGSRVPHIV	RKDLTKYTMM	DLVYALRHFD	RNDCLMLCDI	LSIYAGCEQS	-----YFTKKD
	MHV	VEGSRVPHIV	RKDLTKYTMM	DLVYALRHFD	RNDCLMLCDI	LSIYAGCEQS	-----YFTKKD
	AIPV	KN---IYNIS	RQRLTKYTMM	DFCYALRHFD	PKDCVLEKEI	LVTYGCIEDY	HPKWFEENKD
	SARS CoV	VDGDMVPHIS	RQRLTKYTMA	DLVYALRHFD	EGNCDTLKEI	LVTYNCCDD	-----YFNKKD

		.... ....  .... ....  .... ....  .... ....  .... ....
		185 195 205 215 225 235
5	EMCR	WYDFVENEDI HRVYASLGKI VARAMLKCVA LCDAMVAKGV VGVLTLDNQD LNGNFYDFGD
	229E	WFDPIENEDI HRVYAALGKV VANAMLKCVA FCDENVLKG VGVLTLDNQD LNGNFYDFGD
	PEDV	WFDPIENEDI HRVYALLGTI VARAMLKCVK FCDAMVEQGI VGVVTLTDNQD LNGDFYDFGD
	TGEV	WFDPVENEAI HEVYAKLGPI VANAMLKCVA FCDALVEKGY IGVITLDNQD LNGNFYDFGD
	BoCoV	WYDFVENPDI INVYKKGPI FNRALVSATE FADKLVEVGL VGILTLDNQD LNGKWDYDFGD
	OC43	WYDFVENPDI INVYKKGPI FNRALVSATE FADKLVEVGL VGVLTLDNQD LNGKWDYDFGD
10	MHV	WYDFVENS DI INVYKKGPI FNRALLNTAK FADTLVEAGL VGVLTLDNQD LYQWYDFGD
	AIPV	WYDPIENSKY YVMLAKMGPI VRRALLNAIE FGNLMVEKGY VGVITLDNQD LNGKFYDFGD
	SARS CoV	WYDFVENPDI LRVYANLGER VRQSLLKTQV FCDAMRDAGI VGVLTLDNQD LNGNWDYDFGD
		.... ....  .... ....  .... ....  .... ....  .... ....
		245 255 265 275 285 295
15	EMCR	FVVS LPNMGV PCCTSYYSYM MPIMGLTNCL ASECFVKSDI FGSDFKTFDL LKYDFTEHKE
	229E	FVLCPPGMGI PYCTSYYSYM MPVMGMTNCL ASECFMKSDI FGQDFKTFDL LKYDFTEHKE
	PEDV	FTCSIKMGV PICTSYYSYM MPVMGMTNCL ASECFVKSDI FGEDFKSYDL LEYDFTEHKT
	TGEV	FVKTAGPGCG ACVTSYYSYM MPIMGMTSCL ESENFVKSDI YGSDYKQYDL LAYDFTEHKE
	BoCoV	YVIAAPGCGV AIADSYYSYM MPMLTMCHAL DCELYVNNAY R-----LFDL VQYDFTDYKL
	OC43	YVIAAPGCGV AIADSYYSYM MPMLTMCHAL DCELYVNNAY R-----LFDL VQYDFTDYKL
	MHV	FVKTVPGCGV AVADSYYSYM MPMLTMCHAL DSELFINGTY R-----EFDL VQYDFTDYKL
	AIPV	FQKTAPGAGV PVFDYYSYM MPIIAMTDAL APERYFEYDV HKG-YKSYDL LKYDYTEEKQ
	SARS CoV	FVQVAPGCGV PIVDSYYSLL MPILTLTRAL AAESHMDADL AKP-LIKWDL LKYDFTEERL
		.... ....  .... ....  .... ....  .... ....  .... ....
		305 315 325 335 345 355
30	EMCR	NLFNKYFKHW SFDYHPNCSD CYDDMCVHIC ANFNITLFATT IPGTAFGPLC RKVFDIGVPL
	229E	VLFNKYFKYW GDYHPDCVD CHDEMCIHLC SNFNITLFATT IPNTAFGPLC RKVFDIGVPL
	PEDV	ALFNKYFKYW GLQYHPNCVD CSDEQCIHLC ANFNITLFSTT IPITAFGPLC RKVFDIGVPL
	TGEV	YLFQYFKYW DRTYHPNCSD CTSDECIHLC ANFNITLFSTT IPMTAFGPLV RKVFDIGVPL
	BoCoV	ELFNKYFKHW SMPYHPNTVD CQDDRCIIHC ANFNILFSMV LPNTCFGPLV RQIFVDGVFP
	OC43	ELFNKYFKHW SMPYHPNTVD CQDDRCIIHC ANFNILFSMV LPNTCFGPLV RQIFVDGVFP
	MHV	ELFNKYFKYW SMTYHPNTCE CEDDRCIIHC ANFNILFSMV LPKTCFGPLV RQIFVDGVFP
35	AIPV	ELFQYFKYW DQYHPNCRD CSDDRCIIHC ANFNILFSTL IPQTSFGNLC RKVFDIGVFP
	SARS CoV	CLFDRYFKYW DQYHPNCIN CLDDRCIIHC ANFNILFSTV FPPTSGGPLV RKIFVDGVFP
		.... ....  .... ....  .... ....  .... ....  .... ....
		365 375 385 395 405 415
40	EMCR	VTTAGYHFQ LGLVWNKD VN THSVRLTITE LLQFVTDPSL IASSPALVD QRTICFSVAA
	229E	VATAGYHFQ LGLVWNKD VN THSTRLTITE LLQFVTDPTL IVASSPALVD KRTVCFVSAA
	PEDV	VTTAGYHFQ LGIVWNNDLN LHSSRLSINE LLQFCSDPAL LIASSPALVD QRTVCFVSAA
	TGEV	VVTAGYHFQ LGIVWNLDVK LDTMKLSMTD LLRFVTDPTL LVASSPALD QRTVCFVSAA
	BoCoV	VVSIGYHYKE LGIVMNMDVD THRYRLSLKD LLLYAADPAL HVASASALYD LRTCCFSVAA
	OC43	VVSIGYHYKE LGIVMNMDVD THRYRLSLKD LLLYAADPAL HVASASALYD LRTCCFSVAA
	MHV	VVSIGYHYKE LGVVMNMDVD THRYRLSLKD LLLYAADPAL HVASASALLD LRTCCFSVAA
	AIPV	IATCGYHSKE LGVIMNQDNT MSFSKMGSLQ LMQFVGDPAL LVGTSNNLVD LRTSCFSVCA
	SARS CoV	VVSTGYHFRE LGVVHNQDVN LHSSRLSFKE LLVYAADPAM HAASGNLLD KRTTCFSVAA
		.... ....  .... ....  .... ....  .... ....  .... ....
		425 435 445 455 465 475
50	EMCR	LSTGLTNQV KPGHFNKEFY NFLRLRGFFD EGSELT LKHF FFAQNGDAV KDFDFYRYNK
	229E	LSTGLTSQTV KPGHFNKEFY DFLRSQGFDD EGSELT LKHF FFTQKGDAI KDFDYRYNR
	PEDV	LGTGMTNQTV KPGHFNKEFY DFLRQGFDS EGSELT LKHF FFAQKVDAV KDFDYRYNR
	TGEV	LSTGITYQTV KPGHFNKDFY DFTITERGFFE EGSELT LKHF FFAQGGGAAM TDFNYYRYNR
	BoCoV	ITSGVKFQTV KPGNFNQDFY DFILSKGLLK EGSSVDLKH FTTQDGNAAI TDYNYKYNL
	OC43	ITSGVKFQTV KPGNFNQDFY DFILSKGLLK EGSSVDLKH FTTQDGNAAI TDYNYKYNL
	MHV	ITSGVKFQTV KPGNFNQDFY DFILSKGLLK EGSSVDLKH FTTQDGNAAI TDYNYKYNL
	AIPV	LTSGITHQTV KPGHFNKDFY DFAEKAGMEK EGSSIPLKH FYPQTGNAAI NDYDYRYNR
60	SARS CoV	LTNNVAFQTV KPGNFNKDFY DFAVSKGFFK EGSSVELKH FFAQDGNAAI SDYDYRYNL
		.... ....  .... ....  .... ....  .... ....  .... ....
		485 495 505 515 525 535
65	EMCR	PTILDICQAR VTYKIVSRYF DIYEGGCICA CEVVVTNLNK SAGWPLNKFG KASLYYESIS
	229E	PTMLDIGQAR VAYQVAARYF DCYEGGCITS REVVVTNLNK SAGWPLNKFG KAGLYYESIS
	PEDV	PTVLDICQAR VVYQIVQRYF DIYEGGCITA KEVVVTNLNK SAGYPLNKFG KAGLYYESLS
	TGEV	VTVLDICQAR FVYKIVGKYF ECDYGGCINA REVVVTNYDK SAGYPLNKFG KARLYYETLS
	BoCoV	PTMVDIKQLL FVLEVVKYF EIYDGGCIPA AQVIVNNYDK SAGYPFNKFG KARLYYEALS
	OC43	PTMVDIKQLL FVLEVVKYF EIYDGGCIPA SQVIVNNYDK SAGYPFNKFG KARLYYEALS
	MHV	PTMVDIKQLL FVLEVNNKYF EIYDGGCIPA TQVIVNNYDK SAGYPFNKFG KARLYYEALS
	AIPV	PTMFIDICQLL FCLEVTSKYF ECDYGGCIPA SQVIVNNLDK SAGYPFNKFG KARLYYE-MS
	SARS CoV	PTMCDIRQLL FVVEVVDKYF DCYDGGCINA NQVIVNNLDK SAGFFPNKFG KARLYYDSMS
		.... ....  .... ....  .... ....  .... ....  .... ....
		545 555 565 575 585 595
75	EMCR	YEEQDALFAL TKNRVLPMT QNLNKYAISG KERARTVGGV SLLSTMTTRQ YHOKHLKSIV
	229E	YEEQDAIFSL TKNRILPTMT QNLNKYAISG KERARTVGGV SLLATMTTRQ FHQCLKSIV
	PEDV	YEEQDELYAY TKNRILPTMT QNLNKYAISG KERARTVGGV SLLSTMTTRQ YHOKHLKSIV
	TGEV	YEEQDALFAL TKNRVLPMT QMNLYAISG KARARTVGGV SLLSTMTTRQ YHOKHLKSIA
80	BoCoV	FEEQDEIYAY TKNRVLPMT QMNLYAISA KNRARTVAGV SILSTMTGRM FHQCLKSIA

5	OC43	FEEQDEIYAY	TKRNVLP TLT	QMNLYKAI SA	KNRARTVAGV	SILSTMTGRM	FHQKCLKSIA
	MHV	FEEQDEVYAY	TKRNVLP TLT	QMNLYKAI SA	KNRARTVAGV	SILSTMTGRM	FHQKCLKSIA
	AIPV	LEEQDQLF EI	TKKNVLP TIT	QMNLYKAI SA	KNRARTVAGV	SILSTMTNRQ	FHQKILKSIV
	SARS CoV	YEDQDALFAY	TKRNVIP TIT	QMNLYKAI SA	KNRARTVAGV	SICSTMTNRQ	FHQKLLKSIA
		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
10	EMCR	605	615	625	635	645	655
	229E	NTRNATVVIG	TTKFYGGWNN	MLRTLIDGVE	NPMLMGWDYP	KCDRALPNMI	RMISAMVLGS
	PEDV	ATRNATVVIG	TTKFYGGWDN	MLKNLMADVD	DPKLMGWDYP	KCDRAMPSMI	RMLSAMILGS
	TGEV	NTRGASVVIG	TTKFYGGWDN	MLKNLIDGVE	NPCLMGWDYP	KCDRALPNMI	RMISAMILGS
	BoCoV	ATRGVPVVIG	TTKFYGGWDD	MLRRLIKDVD	NPVLMGWDYP	KCDRAMPNIL	RIVSSSLVLAR
15	OC43	ATRGVPVVIG	TTKFYGGWDD	MLRRLIKDVD	NPVLMGWDYP	KCDRAMPNIL	RIVSSSLVLAR
	MHV	ATRGVPVVIG	TTKFYGGWDD	MLRRLIKDVD	SPVLMGWDYP	KCDRAMPNIL	RIASSSLVLAR
	AIPV	NTRNASVVIG	TTKFYGGWDN	MLRNLIOGVE	DPILMGWDYP	KCDRAMPNLL	RIAASLVLAR
	SARS CoV	ATRGATVVIG	TSKFYGGWHN	MLKTVYS DVE	TPHLMGWDYP	KCDRAMPNML	RIMASLVLAR
		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
20	EMCR	665	675	685	695	705	715
	229E	KHVNCCTVTD	RFYRLGNE LA	QVLTEVVYSN	GGFYFKPGGT	TSGDASTAYA	NSIFNI FOAV
	PEDV	KHVTCC TASD	KFYRLSNE LA	QVLTEVVYSN	GGFYFKPGGT	TSGDATTAYA	NSVFNI FOAV
	TGEV	KHTTCCSSTD	RFFRLCNE LA	QVLTEVVYSN	GGFYFKPGGT	TSGDATTAYA	NSVFNI FOAV
	BoCoV	KHVGCC THND	RFYRLSNE LA	QVLTEVVHCT	GGFYFKPGGT	TSGDATTAYA	NSAFNI FOAV
25	OC43	KHEACCSQSD	RFYRLANE CA	QVLSEIVMCG	GCYVVKPGGT	SSGDATTAF A	NSVENICQAV
	MHV	KHETCCSQSD	RFYRLANE CA	QVLSEIVMCG	GCYVVKPGGT	SSGDATTAF A	NSVENICQAV
	AIPV	KHDSCCSHTD	RFYRLANE CA	QVLSEIVMCG	GCYVVKPGGT	SSGDATTAF A	NSVENICQAV
	SARS CoV	KHTNCCSWSE	RIYRLYNE CA	QVLSETVLAT	GGIYVKPGGT	SSGDATTAF A	NSVFNI IQAT
		KHNTCCNL SH	RFYRLANE CA	QVLSEVMVCG	GSLYVKPGGT	SSGDATTAF A	NSVFNI IQAT
30	EMCR	725	735	745	755	765	775
	229E	SSNINRLLSV	PSDSCNNVNV	RDLQRRLYDN	CYRLTSVEES	FIDDDYGYLR	KHFSMMILSD
	PEDV	SSNINCVLSV	NSSNCCNNFV	KKLQORQLYDN	CYRNSNVDES	FVDDFYGYLQ	KHFSMMILSD
	TGEV	SANVNKLLSV	DSNVCHNLEV	KQLQRKLYEC	CYRSTIVDDQ	FVVEYGYLR	KHFSMMILSD
	BoCoV	SANVNKLLGV	DSNACNNVTV	KSIQRKIYDN	CYRSSSIDEE	FVVEYFSLR	KHFSMMILSD
35	OC43	SANVCALMSC	NGNKIEDLSI	RALQKRLYSH	VYRSDMV DST	FVTEYEF LN	KHFSMMILSD
	MHV	SANVCALMSC	NGNKIEDLSI	RALQKRLYSH	VYRSDKVDST	FVTEYEF LN	KHFSMMILSD
	AIPV	SANVARLLSV	ITRDIVYDNI	KSLQYELYQQ	VYRRVNF DPA	FVEKFYSYLC	KNFSLMILSD
	SARS CoV	TANVNALLST	DGNKIADKYV	RNLQHRLYEC	LYRNRDVDHE	FVDEFYAYLR	KHFSMMILSD
		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
40	EMCR	785	795	805	815	825	835
	229E	DGVVCYNKDY	AELGYIADIS	AFKATLYYQN	NVFMSTSKCW	VEEDLTGKPH	EFCSQHTMQI
	PEDV	DSVVCYNKTY	AGLGYIADIS	AFKATLYYQN	GVFMSTAKCW	TEEDLSIGPH	EFCSQHTMQI
	TGEV	DGVVCYNNDY	ASLG YVADLN	AFKAVLYYQN	NVFMASASKCW	IEPDINKGPH	EFCSQHTMQI
	BoCoV	DGVVCYNKDY	ADLGYVADIN	AFKATLYYQN	NVFMSTSKCW	VEPDLSVGPH	EFCSQHTLQI
45	OC43	DGVVCYNNDY	ASKGYIANIS	AFQOQVLYQN	NVFMSESKCW	VENDINNGPH	EFCSQHTMLV
	MHV	DGVVCYNNDY	ASKGYIANIS	AFQOQVLYQN	NVFMSESKCW	VEHDINNGPH	EFCSQHTMLV
	AIPV	DGVVCYNSEF	ASKGYIANIS	AFQOQVLYQN	NVFMSEAKCW	VETDIEKGPH	EFCSQHTMLV
	SARS CoV	DGVVCYNNTL	AKQGLVADIS	GFREVLYYQN	NVFMADSKCW	VEPDLEKGPH	EFCSQHTMLV
		DAVVCYN SNY	AAQGLVASIK	NFKAVLYYQN	NVFMSEAKCW	TETDLTKGPH	EFCSQHTMLV
50	EMCR	845	855	865	875	885	895
	229E	VDKDGTYYLP	YDPDSRILSA	GVFVDDVVK T	DAVLLXRYV	SLAIDAYPLS	KHPNSEYRKV
	PEDV	VDENGKYYLP	YDPDSRIISA	GVFVDDITKT	DAVILLERYV	SLAIDAYPLS	KHPKPEYRKV
	TGEV	VDKEGTYYLP	YDPDSRILSA	GVFVDDVVK T	DAVLLERYV	SLAIDAYPLS	KHPKPEYRKV
	BoCoV	VGPDGDYYLP	YDPDSRILSA	GVFVDDIVKT	DNVIMLERYV	SLAIDAYPLT	KHPKPAYQKV
55	OC43	KMDGDDVYLP	YPVPSRILGA	GCFVDDLKLT	DSVLLIERFV	SLAIDAYPLV	YHENEYQKV
	MHV	KMDGDDVYLP	YPNPSRILGA	GCFVDDLKLT	DSVLLIERFV	SLAIDAYPLV	YHENEYQKV
	AIPV	KMDGDEVYLP	YDPDSRILGA	GCFVDDLKLT	DSVLLIERFV	SLAIDAYPLV	YHENEYQKV
	SARS CoV	EVDGEPKYLP	YDPDSRILGA	CVFVDDVDKT	EPVAVMERYI	ALADAYPLV	HHENEYQKV
		KQGDYVYLP	YDPDSRILGA	GCFVDDIVKT	DGTLMIERFV	SLAIDAYPLT	KHPNQEYADV
60	EMCR	905	915	925	935	945	955
	229E	FYVLLDWVKH	LNKNLNEGVL	ESFSVTLLDN	QEDKEWCEDE	YASMYENSTI	LQAAGLCVVC
	PEDV	FYVLLDWVKH	LNKNLNEGVL	ESFSVTLLDE	HESKFWDSEF	YASMYEKSTV	LQAAGLCVVC
	TGEV	FYVLLDWVKH	LYKTLNAGVL	ESFSVTLLDE	STAKFWDSEF	YANMYEKS AV	LQAAGLCVVC
	BoCoV	FYVLLDWVKH	LQKNL NAGVL	DSFSVTMLLE	GQDKFWSEEF	YASLYEKSTV	LQAAGLCVVC
65	OC43	FRVYLEYIKK	LYNELGNQIL	DSYSVILSTC	DGQKFTDESE	YKNMYLRS AV	MQSVGACVVC
	MHV	FRVYLEYIKK	LYNDLGNQIL	DSYSVILSTC	DGQKFTDESE	YKNMYLRS AV	MQSVGACVVC
	AIPV	FRVYLEYIKK	LYNDLGNQIL	DSYSVILSTC	DGQKFTDETF	YKNMYLRS AV	MQSVGACVVC
	SARS CoV	FFVLLAYIRK	LYQELSONML	MDYSFVMDID	KGSKFWEQEF	YENMYRAPTT	LQSCGVCVVC
		FHLYLQYIRK	LHDELTGHML	DMYSVMLTND	NTSRYWEPEE	YEAMYPHTV	LQAVGACVLC
70	EMCR	965	975	985	995	1005	1015
	229E	GSQTVLRCGD	CLRKPM LCTK	CAYDHVFGTD	HKEFILAITPY	VCNASGCGVS	DVKKLYLGLL
	PEDV						
	TGEV						
	BoCoV						
75	OC43						
	MHV						
	AIPV						
	SARS CoV						
80	EMCR						

5	229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	GSQTVLRCGD GSQTVLRCGD GSQTVLRCGD SSQTSRLRCGS SSQTSRLRCGS SSQTSRLRCGS NSQTLILRCGN NSQTSRLRCGA	CLRRPMLCTK CLRRPMLCTK CLRRPMLCTK CIRKPLLCKK CIRKPLLCKK CIRKPLLCKK CIRKPLLCKK CIRKPLLCKK	CAYDHVFGTD CAYDHVIGTT CAYDHVMGTT CCYDHVMATD CCYDHVMATD CAYDHVMSTD CCYDHVMHTD CCYDHVISTS	HKFILAITPY HKFILAITPY HKFILAITPY HKYVLSVSPY HKYVLSVSPY HKYVLSVSPY HKNVLSINPY HKLVLVSNPY	VCNTSGCNVN VCCASDCGVN VCSFNGCNVN VCNAPGCDVN VCNAPGCDVN VCNSPGCDVN ICSQLGCGEA VCNAPGCDVT	DVTKLYLGGL DVTKLYLGGL DVTKLYLGGL DVTKLYLGGM DVTKLYLGGM DVTKLYLGGM DVTKLYLGGM DVTQLYLGGM
10		.... ....  1025	.... ....  1035	.... ....  1045	.... ....  1055	.... ....  1065	.... ....  1075
15	EMCR 229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	NYCYTNHKKPQ NYCYVDHKKPH SYWCHHKKPR SYCYMNHKKPQ SYCYEDHKKPQ SYCYEDHKKPQ SYFCGNHKKPK SYCKSHKPP	LSFPLCSAGN LSFPLCSAGN LAFFPLCSAGN LSFPLCANGN YSFKLVNMG YSFKLVNMG LSIPLVSNGT ISFPLCANGQ	IFGLYKNSAT VFGLYKSSAL VFGLYKNSAT VFGLYKSSAV VFGLYKQSC VFGLYKQSC VFGLYKQSC VFGLYKNTCV	GSLDVEVFNR GSMDIDVFNK GSPDVEDFNR GSEAVEDFNK GSPYIDDFNR GSPYIDDFNR GSPYIDDFNR GSENVDDFNQ	LATSDWTDVR LSTDWSDIR IATSDWTDVS LAVSDWTNVE IASCKWTDVD IASCKWTDVD IASCKWTEVD LATTNWSIVE	DYKLANDVKD DYKLANDAKE DYRLANDVKD DYKLANNVKE DYILANECTE DYILANECTE DYVLANECTE PYILANRCS
20		.... ....  1085	.... ....  1095	.... ....  1105	.... ....  1115	.... ....  1125	.... ....  1135
25	EMCR 229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	TLRLFAAETI SLRLFAAETV SLRLFAAETI SLKIFAAETV RLKLFAAETQ RLKLFAAETQ RLKLFAAETQ SLRRFAAETV RLKLFAAETL	KAKEESVKSS KAKEESVKSS KAKEESVKSS KAKEESVKSE KATEEAFKQS KATEEAFKQS KATEESFKQC KATEELHKQQ KATEETFKLS	YAFATLKEVV YAYATLKEIV YACATLHEVV YAYAVLKEVI YASATIQEIV YASATIQEIV YASATIREIV FASAEVREVF YGIATVREVL	GPKELLSWE GPKELLLLWE GPKELLLKWE GPKELIVLQWE SERELILSWE SERELILSWE SDRELILSWE SDRELILSWE SDRELHLSWE	SGKVKPPLNR SGKAKPPLNR VGRPKPPLNR ASKTKPPLNR IGKVKPPLNK IGKVKPPLNK IGKVRPPLNK PGKTRPPLNR VGKPRPPLNR	NSVFTCFQIS NSVFTCFQIT NSVFTCYHIT NSVFTCFQIS NYVFTGYHFT NYVFTGYHFT NYVFTGYHFT NYVFTGYHFT NYVFTGYRVT
30		.... ....  1145	.... ....  1155	.... ....  1165	.... ....  1175	.... ....  1185	.... ....  1195
35	EMCR 229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	KDSKFQIGEF KDSKFQVGEF KNTKFQIGEF KDTKIQIGEF KNGKTVLGEY KNGKTVLGEY SNGKTVLGEY RTSKVQLGDF KNSKVQIGEX	IFEKVEYGS VFEKVDYGS VFEKAEYDND VFEQSEYGS VFDKSEL-TN VFDKSEL-TN VFDKSEL-TN TFEKGEG-KD TFEKGDY-GD	TVTYKSTVTT TVTYKSTATT AVTYKTATT SVYYKSTSTY GVYYRATTTY GVYYRATTTY GVYYRATTTY VVYYKATSTA AVVYRGTTY	KLVPGMIFVL KLVPGMIFIL KLVPGMVFVL KLTPGMIFVL KLSVGDVFL KLSVGDVFL KLSVGDVFL KLSVGDIFVL KLVNGDYFVL	TSHNVQPLRA TSHNVAPLRA TSHNVQPLRA TSHNVSPKLA TSHSVANLSA TSHSVANLSA TSHAVSSLSA TSHNVVSLVA TSHTVMPLSA	PTIANQEKYS PTMANQEKYS PTIANQERY PILVNQEKYN PTLVPQENYS PTLVPQENYS PTLVPQENYT PTLCPQQTFS PTLVPQEHYV
40		.... ....  1205	.... ....  1215	.... ....  1225	.... ....  1235	.... ....  1245	.... ....  1255
45	EMCR 229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	SIYKLHPAFN TIYKLHPSPFN TIHKLHPAFN TISKLYPVFN SIR-FASVYS SIR-FASVYS SIR-FASVYS RVNLRPNVM RITGLYPTLN	VSDAYANLVP VSDAYANLVP IPEAYSSLVP IAEAYNTLVP VLETQNNV VLETQNNVV VPETEQQNV VPECFFVNIP ISDEFSSNVA	YYQLIGKQKI YYQLIGKQRI YYQLIGKQKI YYQMIGKQKF NYQHIGMKRY NYQHIGMKRY NYQHIGMKRY LYHLVKGQKR NYQVGMQKY	TTIQGPPGSG TTIQGPPGSG TTIQGPPGSG TTIQGPPGSG CTVQGPPTG CTVQGPPTG CTVQGPPTG TTVQGPPTG STLQGPPTG	KSHCSIGLGL KSHCSIGIGV KSHCVIGLGL KSHCVIGLGL KSHLAIGLAV KSHLAIGLAV KSHLAIGLAV KSHFAIGLAV KSHFAIGLAL	YYPGARIVFV YYPGARIVFT YYPGARIVFT YYPGARIVFT YFCTARVVYT YFCTARVVYT YFCTARVVYT YFSSARVVFT YYPGARIVFT
50		.... ....  1265	.... ....  1275	.... ....  1285	.... ....  1295	.... ....  1305	.... ....  1315
55	EMCR 229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	ACSHAAVDSL ACSHAAVDSL ACSHAAVDSL ACSHAAVDAL AASHAAVDAL AASHAAVDAL AASHAAVDAL ACSHAAVDAL ACSHAAVDAL	CAKAMTVYSI CAKAVTAYSV CVKASTAYSN CEKAAKNFNV CEKAYKFLNI CEKAYKFLNI CEKAYKFLNI CEKAFKFLKV CEKALKYLP	DKCTRIIPAR DKCTRIIPAR DKCSRIIPQR DRCSRIIPQR NDCTRIIPAK NDCTRIIPAK NDCTRIIPAK DDCTRIIPQR DKCSRIIPAR	ARVECYSGFK ARVECYSGFK ARVECYDGFK IRVDCYTGFK VRVECYDKFK VRVECYDKFK VRVDCYDKFK TTVDCFSKFK ARVECFDKFK	PNNNTSAQYIF PNNNTSAQYVF PNNNTSAQYLF PNNNTAQYLF INDTTRKYVF INDTTRKYVF VNDTTRKYVF ANDTGKKYIF VNSTLEQYVF	STVNALPECN STVNALPEVN STVNALPECN CTVNALPEAS TTINALPEMV TTINALPEMV TTINALPELV TTINALPEVS CTVNALPETT
60		.... ....  1325	.... ....  1335	.... ....  1345	.... ....  1355	.... ....  1365	.... ....  1375
65	EMCR 229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	ADIVVVDEVS ADIVVVDEVS ADIVVVDEVS CDIVVVDEVS TDIVVVDEVS TDIVVVDEVS TDIVVVDEVS CDILLVDEVS ADIVVFDEIS	MCTNYDLSVI MCTNYDLSVI MCTNYDLSVI MCTNYDLSVI MLTNYELSVI MLTNYELSVI MLTNYELSVI MLTNYELSVI MATNYDLSV	NQRLSYKHIV NQRLSYKHIV NQRLSYKHIV NSRLSYKHIV NARIRAKHYV NARIRAKHYV NSRVRAKHIV NGKINYQYVV NARIRAKHYV	YVGDPQQLPA YVGDPQQLPA YVGDPQQLPA YVGDPQQLPA YIGDPAQLPA YIGDPAQLPA YIGDPAQLPA YIGDPAQLPA YIGDPAQLPA	PRVMITRGVM PRVLISRGVM PRVMISRGTL PRTLINRGVL PRVLLSKGTL PRVLLSKGTL PRVLLSKGTL PRTLLN-GSL PRTLLTKGTL	EPVDYNVVTQ EPIDYNVVTQ EPKDYNNVVTQ QPDYNNVVTQ EPKYFNTVTK EPKYFNTVTK EPKYFNTVTK SPKDYNNVVT EPKYFNTVCR
70		.... ....  1385	.... ....  1395	.... ....  1405	.... ....  1415	.... ....  1425	.... ....  1435
75	EMCR 229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	NYCYTNHKKPQ NYCYVDHKKPH SYWCHHKKPR SYCYMNHKKPQ SYCYEDHKKPQ SYCYEDHKKPQ SYFCGNHKKPK SYCKSHKPP	LSFPLCSAGN LSFPLCSAGN LAFFPLCSAGN LSFPLCANGN YSFKLVNMG YSFKLVNMG LSIPLVSNGT ISFPLCANGQ	IFGLYKNSAT VFGLYKSSAL VFGLYKNSAT VFGLYKSSAV VFGLYKQSC VFGLYKQSC VFGLYKQSC VFGLYKNTCV	GSLDVEVFNR GSMDIDVFNK GSPDVEDFNR GSEAVEDFNK GSPYIDDFNR GSPYIDDFNR GSPYIDDFNR GSENVDDFNQ	LATSDWTDVR LSTDWSDIR IATSDWTDVS LAVSDWTNVE IASCKWTDVD IASCKWTDVD IASCKWTEVD LATTNWSIVE	DYKLANDVKD DYKLANDAKE DYRLANDVKD DYKLANNVKE DYILANECTE DYILANECTE DYVLANECTE PYILANRCS
80		.... ....  1445	.... ....  1455	.... ....  1465	.... ....  1475	.... ....  1485	.... ....  1495
85	EMCR 229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	TLRLFAAETI SLRLFAAETV SLRLFAAETI SLKIFAAETV RLKLFAAETQ RLKLFAAETQ RLKLFAAETQ SLRRFAAETV RLKLFAAETL	KAKEESVKSS KAKEESVKSS KAKEESVKSS KAKEESVKSE KATEEAFKQS KATEEAFKQS KATEESFKQC KATEELHKQQ KATEETFKLS	YAFATLKEVV YAYATLKEIV YACATLHEVV YAYAVLKEVI YASATIQEIV YASATIQEIV YASATIREIV FASAEVREVF YGIATVREVL	GPKELLSWE GPKELLLLWE GPKELLLKWE GPKELIVLQWE SERELILSWE SERELILSWE SDRELILSWE SDRELILSWE SDRELHLSWE	SGKVKPPLNR SGKAKPPLNR VGRPKPPLNR ASKTKPPLNR IGKVKPPLNK IGKVKPPLNK IGKVRPPLNK PGKTRPPLNR VGKPRPPLNR	NSVFTCFQIS NSVFTCFQIT NSVFTCYHIT NSVFTCFQIS NYVFTGYHFT NYVFTGYHFT NYVFTGYHFT NYVFTGYHFT NYVFTGYRVT



5	EMCR	1385	1395	1405	1415	1425	1435
	229E	RMCAIGPDVF	LHKCYRCPAE	IVNTVSELVY	ENKFVPVKPA	SKQCFKIFFK	G---NVQVDN
	PEDV	RMCAIGPDVF	LHKCYRCPAE	IVNTVSELVY	ENKFVPVKPA	SKQCFKIFER	G---SVQVDN
	TGEV	RMCTLGPDVF	LHKCYRCPAE	IVRTVSEMVY	ENQFIPVHPD	SKQCFKIFCK	G---NVQVDN
	BoCoV	LMCCLGPDIF	LGTCTYRCPKE	IVDTVSALVY	ENKLFVPVNPPE	SKQCFKMFVK	G---QVQIES
10	OC43	LMCCLGPDIF	LGTCTYRCPKE	IVDTVSALVY	ENKLFVPVNPPE	SKQCFKMFVK	G---VTTHES
	MHV	LMCCLGPDIF	LGTCTYRCPKE	IVDTVSALVY	ENKLFVPVNPPE	SKQCFKMFVK	G---VTTHES
	AIPV	LMVCVKPDIF	LAKCYRCPKE	IVDTVSTLVY	DNKLFVANNPE	SSMCFKVVYK	G---VTTHES
	SARS	LMKTIGPDMF	LGTCTYRCPKE	IVDTVSALVY	DNKLFVANNPE	SSMCFKVVYK	G---VTTHES
	CoV	LMKTIGPDMF	LGTCTYRCPKE	IVDTVSALVY	DNKLFVANNPE	SSMCFKVVYK	G---VTTHES
15	EMCR	1445	1455	1465	1475	1485	1495
	229E	GSSINRRQLE	IVKLFVLKPN	SWSKAVFISP	YNSQNYVASR	FLGLQIQTV	SSQSEYDYV
	PEDV	GSSINRRQLE	IVKLFVLKPN	SWSKAVFISP	YNSQNYVASR	FLGLQIQTV	SSQSEYDYV
	TGEV	GSSINRRQLE	IVKLFVLKPN	SWSKAVFISP	YNSQNYVASR	FLGLQIQTV	SSQSEYDYV
	BoCoV	SSAVNMQQIY	LINKFLKANP	LWHKAVFISP	YNSQNYFAAKR	VLGLQIQTV	SSQSEYDYV
20	OC43	SSAVNMQQIY	LINKFLKANP	LWHKAVFISP	YNSQNYFAAKR	VLGLQIQTV	SSQSEYDYV
	MHV	SSAVNMQQIY	LINKFLKANP	LWHKAVFISP	YNSQNYFAAKR	VLGLQIQTV	SSQSEYDYV
	AIPV	GSAYNTTQLE	FVKDFVCRNK	AWRKAVFISP	YNSQNYFAAKR	VLGLQIQTV	SSQSEYDYV
	SARS	SSAINRPQIG	VVREFLTRNP	AWRKAVFISP	YNSQNYFAAKR	VLGLQIQTV	SSQSEYDYV
	CoV	SSAINRPQIG	VVREFLTRNP	AWRKAVFISP	YNSQNYFAAKR	VLGLQIQTV	SSQSEYDYV
25	EMCR	1505	1515	1525	1535	1545	1555
	229E	IYAQTSDDTAH	ACNVNRFNVA	ITRAKKGIFC	VMCDKT-LFD	SLKFFFEIKHA	---DLHSS---
	PEDV	IYAQTSDDTAH	ACNVNRFNVA	ITRAKKGIFC	VMCDKT-LFD	SLKFFFEIKHA	---DLHSS---
	TGEV	IYAQTSDDTAH	ACNVNRFNVA	ITRAKKGIFC	VMCDKT-LFD	SLKFFFEIKHA	---DLHSS---
	BoCoV	IYTQTSDDTAH	ATNVNRFNVA	ITRAKKGILC	IMCDRS-LFD	LLKFFFEIKLS	---DLQAN---
30	OC43	IYTQTSDDTAH	ATNVNRFNVA	ITRAKKGILC	IMCDRS-LFD	LLKFFFEIKLS	---DLQAN---
	MHV	IYSQTAETAH	SVNVNRFNVA	ITRAKKGILC	IMCDRT-MYE	NLDFFELKDS	KIGLOAKEP-
	AIPV	IYSQTAETAH	SVNVNRFNVA	ITRAKKGILC	IMCDRT-MYE	NLDFFELKDS	KIGLOAKEP-
	SARS	IYSQTAETAH	SVNVNRFNVA	ITRAKKGILC	IMCDRT-MYE	NLDFFELKDS	KIGLOAKEP-
	CoV	IYSQTAETAH	SVNVNRFNVA	ITRAKKGILC	IMCDRT-MYE	NLDFFELKDS	KIGLOAKEP-
35	EMCR	1565	1575	1585	1595	1605	1615
	229E	-QVCGLEFKNC	TRTFLNLPPT	HAHTFLSLSD	QFKTTGDLAV	QIGSNN--VC	TYEHVISFMG
	PEDV	-SSCGLEFKDC	ARNPIDLPPS	HATTYLSLSD	RFKTSGDLAV	QIGSNN--VC	TYEHVISFMG
	TGEV	-EGCGLEFKDC	SRGDDLLPPS	HANTFMSLAD	NFKTDQYLAV	QIGVNG--PI	KYEHVISFMG
	BoCoV	-ETCGLEFKDC	SKSEQYIPEA	YATTYMSLSD	NFKTSDGLAV	NIG-TK--DV	KYANVISFMG
40	OC43	QCSTNLFKDC	SKSYSGYHFA	HAPSFLAVDD	KYKATGDLAV	CLGIGD-SAV	TYSRLISLMG
	MHV	QCSTNLFKDC	SKSYSGYHFA	HAPSFLAVDD	KYKATGDLAV	CLGIGD-SAV	TYSRLISLMG
	AIPV	QCTTNLFKDC	SRSYAGYHFA	HAPSFLAVDD	KYKATGDLAV	CLGIGD-SAV	TYSRLISLMG
	SARS	---TGLEFKIC	NKEFGSVHFA	YAVTTKALAA	TYKVNDLAA	LVNVEAGSEI	TYKHLISLLG
	CoV	ENVVTGLEFKDC	SKIITGLHPT	QAPTHLSVDI	KFKTEG-LCV	DIPGIP-KDM	TYRRLISMMG
45	EMCR	1625	1635	1645	1655	1665	1675
	229E	FRFDISIPGS	HSLFCTRDEFA	IRNVRGWLGM	DVEGAHVCGD	NIGTNVPLQV	GFSNGVNFVV
	PEDV	FRFDISIPGS	HSLFCTRDEFA	IRNVRGWLGM	DVEGAHVCGD	NIGTNVPLQV	GFSNGVNFVV
	TGEV	FRFDINIPNH	HTLFCTRDEFA	MRNVRGWLGF	DVEGAHVCGD	NIGTNVPLQV	GFSNGVNFVV
	BoCoV	FRFEANIPGY	HTLFCTRDEFA	MRNVRAWLGF	DVEGAHVCGD	NIGTNVPLQV	GFSNGVNFVV
50	OC43	FKLDVTLDDY	CKLFITKEEA	VKRVRWVGF	DAEGAHAATRD	SIGTNFPLQL	GFSNGVNFVV
	MHV	FKLDVTLDDY	CKLFITKEEA	VKRVRWVGF	DAEGAHAATRD	SIGTNFPLQL	GFSNGVNFVV
	AIPV	FKLDLTLDDY	CKLFITRDEA	IRNVRAWVGF	DAEGAHAATRD	SIGTNFPLQL	GFSNGVNFVV
	SARS	FKMSVNVEGC	HNMFITRDEA	IRNVRAWVGF	DAEGAHAATRD	SIGTNFPLQL	GFSNGVNFVV
	CoV	FKMNYQVNGY	PNMFITRDEA	IRNVRAWVGF	DAEGAHAATRD	SIGTNFPLQL	GFSNGVNFVV
55	EMCR	1685	1695	1705	1715	1725	1735
	229E	QTEGCVSTNF	GDVVKPVCAK	SPPGEQFRHL	VFFLRKGQPW	LIVRRRIVQM	ISDYLSNLSD
	PEDV	QTEGCVSTNF	GDVVKPVCAK	SPPGEQFRHL	VFFLRKGQPW	LIVRRRIVQM	ISDYLSNLSD
	TGEV	RPEGCVVTES	GDYIKPVCAK	APPGEQFAHL	VPLLRKGQPW	DVVRKRIVQM	ISDYLSNLSD
	BoCoV	QTEGCVITEK	GNSIEVVKAR	APPGEQFAHL	VPLMRKGQPW	HIVRRRIVQM	VCDYFDGLSD
60	OC43	EATGLFADRD	GYSFKKAVAK	APPGEQFKHL	IPLMTRGQWR	DVVRPRIVQM	FADHLIDLSD
	MHV	EATGLFADRD	GYSFKKAVAK	APPGEQFKHL	IPLMTRGQWR	DVVRPRIVQM	FADHLIDLSD
	AIPV	EATGMFAERD	GYVFKKAVAR	APPGEQFKHL	VPLMSRGQKW	DVVRIRIVQM	LSDHLVDLAD
	SARS	TPEGLVDTSI	GNNFEPVNSK	APPGEQFNHL	RVLFKSAPKW	HVIRPRIVQM	LADNLCNVSD
	CoV	VPTGYVDTEN	NTEFTRVNAK	PPPGDQFKHL	IPLMYKGLPW	NVVRIRIVQM	LSDTLKGLSD
65	EMCR	1745	1755	1765	1775	1785	1795
	229E	ILVFLWAGS	LELTMMRYFV	KIGPIKYCY-	CGNSATCYNS	VSNEYCCFKH	ALGCDYVYNP
	PEDV	ILVFLWAGS	LELTMMRYFV	KIGAVKHCQ-	CGTVATCYNS	VSNDYCCFKH	ALGCDYVYNP
	TGEV	ILVFLWAGS	LELTMMRYFV	KIGPSKSCD-	CGKVATCYNS	ALHTYCCFKH	ALGCDYLYNP
	BoCoV	ILVFLWAGS	LELTMMRYFV	KIGRPQKCE-	CGKSATCYSS	SQSVYACFKH	ALGCDYLYNP
70	OC43	CVVLVTWAAN	FELTCLRYFA	KVGREISCNV	STKRATAYNS	RTGYGCRWRH	SVTCYDLYNP
	MHV	CVVLVTWAAN	FELTCLRYFA	KVGREISCNV	STKRATAYNS	RTGYGCRWRH	SVTCYDLYNP
	AIPV	CVVLVTWAAN	FELTCLRYFA	KVGREISCNV	STKRATAYNS	RTGYGCRWRH	SVTCYDLYNP
	SARS	CVVLVTWAAN	FELTCLRYFA	KVGREISCNV	STKRATAYNS	RTGYGCRWRH	SVTCYDLYNP
	CoV	CVVLVTWAAN	FELTCLRYFA	KVGREISCNV	STKRATAYNS	RTGYGCRWRH	SVTCYDLYNP
75	EMCR	1745	1755	1765	1775	1785	1795
	229E	ILVFLWAGS	LELTMMRYFV	KIGPIKYCY-	CGNSATCYNS	VSNEYCCFKH	ALGCDYVYNP
	PEDV	ILVFLWAGS	LELTMMRYFV	KIGAVKHCQ-	CGTVATCYNS	VSNDYCCFKH	ALGCDYVYNP
	TGEV	ILVFLWAGS	LELTMMRYFV	KIGPSKSCD-	CGKVATCYNS	ALHTYCCFKH	ALGCDYLYNP
	BoCoV	ILVFLWAGS	LELTMMRYFV	KIGRPQKCE-	CGKSATCYSS	SQSVYACFKH	ALGCDYLYNP
80	OC43	CVVLVTWAAN	FELTCLRYFA	KVGREISCNV	STKRATAYNS	RTGYGCRWRH	SVTCYDLYNP
	MHV	CVVLVTWAAN	FELTCLRYFA	KVGREISCNV	STKRATAYNS	RTGYGCRWRH	SVTCYDLYNP
	AIPV	CVVLVTWAAN	FELTCLRYFA	KVGREISCNV	STKRATAYNS	RTGYGCRWRH	SVTCYDLYNP
	SARS	CVVLVTWAAN	FELTCLRYFA	KVGREISCNV	STKRATAYNS	RTGYGCRWRH	SVTCYDLYNP
	CoV	CVVLVTWAAN	FELTCLRYFA	KVGREISCNV	STKRATAYNS	RTGYGCRWRH	SVTCYDLYNP

56/87

OC43	CVVLVTWAAN	FELTCLRYFA	KVGREISCNV	CTKRATVYNS	RTGYGWCWRH	SVTCDYLYNP
MHV	SVVLVTWAAS	FELTCLRYFA	KVGKEVVCVS	CNKRATCFNS	RTGYGWCWRH	SYSCDYLYNP
AIPV	CVVFTWCHG	LELTTLRYFV	KIGKEQVCS-	CGRATTENS	HTQAYACWKH	CLGDFDVYNP
SARS CoV	RVVFLWAHG	FELTSMKYFV	KIGPERTCCL	CDKRATCFST	SSDTYACWNH	SVGFDYVYNP
5	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	1805	1815	1825	1835	1845	1855
EMCR	YAFDIQWGY	VGSLSQNHHT	FCNIHRNEHD	ASGDAVMTRC	LAVHDCFVKN	VDWTVTYPFI
229E	YVIDIQWGY	VGSLSTNHHA	ICNVHRNEHV	ASGDAIMTRC	LAVYDCFVKN	VDWSITYPMI
10 PEDV	YCIDIQWGY	KGSLSLNHHE	HCNVHRNEHV	ASGDAIMTRC	LAIHDCFVKN	VDWSITYPFI
TGEV	YCIDIQWGY	TGSLSMNHHE	VCNIHRNEHV	ASGDAIMTRC	LAIHDCFVKN	VDWSIVYPFI
BoCoV	LIVDIQWGY	IGSLSSNHDL	YCSVHKGAHV	ASSDAIMTRC	LAVYDCFVKN	INWNVEYPII
OC43	LIVDIQWGY	IGSLSSNHDL	YCSVHKGAHV	ASSDAIMTRC	LAVYDCFVKN	INWNVEYPII
MHV	LIVDIQWGY	TGSLTSNHDL	ICSVHKGAHV	ASSDAIMTRC	LAVHDCFVKN	VNWSLEYPII
15 AIPV	LLVDIQWGY	SGNLQFNHDL	HCNVHGHAV	ASVDAIMTRC	LAINNAFCQD	VNWDLTYPHI
SARS CoV	FMIDVQWGF	TGNLQSNHDL	HCQVHGNHVA	ASCDAIMTRC	LAVHECFVKN	VDWSVEYPII
	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	1865	1875	1885	1895	1905	1915
20 EMCR	ANEKFINGCG	RNVQGHVVRA	ALKLYKPSVI	HDIGNPKGVR	CA-VTDAKWY	CYDKQPVNSN
229E	ANENAINKGG	RTVQSHIMRA	AIKLYNPKAI	HDIGNPKGIR	CA-VTDAKWY	CYDKNPINSN
PEDV	GNEAVINKSG	RIVQSHIMRA	VLKLYNPKAI	YDIGNPKGIR	CA-VTDAKWY	CFDKNPTNSN
TGEV	DNEEKINKAG	RIVQSHVMA	ALKIFNPAAI	HDVGNPKGIR	CA-TTPIPF	CYDRDPINNN
BoCoV	SNELISINTSC	RVLQRMVMA	AMLCNRYTLC	YDIGNPKAIA	CV--KDFDFK	FYDAQPIVKS
25 OC43	SNELISINTSC	RVLQRMVMA	AMLCNRYTLC	YDIGNPKAIA	CV--KDFDFK	FYDAQPIVKS
MHV	SNEVSINTSC	RLLQRMVMA	AMLCNRYTLC	YDIGNPKGIR	CV--KGYDFK	FYDASPVVKS
AIPV	ANEDEVNSSC	RYLQRMVMA	CVDALKVNVV	YDIGNPKGIR	CVRGQDVNFR	FYDKNPIVRN
SARS CoV	GDELRVNSAC	RKVQHMVVK	ALLADKEFVL	HDIGNPKAIA	CVPQAEVWK	FYDAQPCSDK
	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	1925	1935	1945	1955	1965	1975
30 EMCR	---VKLLDYD	YATHG---QLD	GLCLFWNCNV	DMYPEFSIVC	RFDTRTRSVF	NLEGVNNGSSL
229E	---VKTLEYD	YATHG---QMD	GLCLFWNCNV	DMYPEFSIVC	RFDTRTRSTL	NLEGVNNGSSL
35 PEDV	---VKTLEYD	YATHG---QMD	GLCLFWNCNV	DMYPEFSIVC	RFDTRTRSTL	NLEGVNNGSSL
TGEV	---VRCLDYD	YATHG---QMD	GLCLFWNCNV	DMYPEFSIVC	RFDTRTRSKL	SLEGCNNGGAL
BoCoV	---VKTLLYF	FEAHKDSFKD	GLCMFWNCNV	DKYPPNAVVC	RFDTRVLNML	NLPGCNNGSSL
OC43	---VKTLLYF	FEAHKDSFKD	GLCMFWNCNV	DKYPPNAVVC	RFDTRVLNML	NLPGCNNGSSL
MHV	---VKQFVYK	FEAHKDSFKD	GLCMFWNCNV	DKYPPNAVVC	RFDTRVLNML	NLPGCNNGSSL
40 AIPV	---VKQFEYD	YNQHKDKFAD	GLCMFWNCNV	DCYPDNLVVC	RYDTRNLVSV	NLPGCNNGSSL
SARS CoV	AYKIEELEYS	YATHHDKETD	GVCLFWNCNV	DRYPANAIVC	RFDTRVLNML	NLPGCDGGSSL
	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	1985	1995	2005	2015	2025	2035
45 EMCR	YVNKHAFHTP	AYDKRAFVKL	KPMPEFFYYDD	SDCDVVQ---	-EQVNYVPLR	ASSCVTRCNI
229E	YVNKHAFHTP	AYDKRAMAKL	KPMPEFFYYDD	GSCEVVH---	-DQVNYVPLR	ATNCITKCN
PEDV	YVNKHAFHTP	AFDKRAFAKL	KPMPEFFYYDD	TECDKLQ---	-DSINYVPLR	ASNCITKCNV
TGEV	YVNKHAFHTP	AYDRAFAKL	KPMPEFFYYDD	SNCELVD---	-GQPNYVPLK	SNVCITKCN
BoCoV	YVNKHAFHTK	PFSRAAFEHL	KPMPEFFYYSD	TPCVYMDGMD	AKQVDYVPLK	SATCITRCNL
50 OC43	YVNKHAFHTK	PFSRAAFEHL	KPMPEFFYYSD	TPCVYMDGMD	AKQVDYVPLK	SATCITRCNL
MHV	YVNKHAFHTS	PFTRAAFENL	KPMPEFFYYSD	TPCVYMEGME	SKQVDYVPLR	SATCITRCNL
AIPV	YVNKHAFYTP	KFDRISEFNL	KAMPEFFYYSD	SPCETIQ-VD	GVAQDLVSLA	TKDCITKCN
SARS CoV	YVNKHAFHTP	AFDKSAFTNL	KQLPFFYYSD	SPCESHGKQV	VSDIDYVPLK	SATCITRCNL
	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	2045	2055	2065	2075	2085	2095
55 EMCR	GGAVCSKHAN	LYQRYVEAYN	TFTQAGFNIW	VPHSFVDVYNL	WQIFET-NL	QSLNIAFNV
229E	GGAVCSKHAN	LYRAYVESYN	IFTQAGFNIW	VPTTFDCYNL	WQTFTEV-NL	QGLNIAFNV
60 PEDV	GGAVCSKHCA	MYHSYVNAYN	TFTSAGFTIW	VPTSFDTYNL	WQTFSN--NL	QGLNIAFNV
TGEV	GGAVCKKHAA	LYRAYVEDYN	IFMQAGFTIW	CPQNFDTYML	WHGFVNSKAL	QSLNIAFNV
BoCoV	GGAVCLKHAE	EYREYLESYN	TATTAGFTFW	VYKTFDFYNL	WNTFTK---L	QSLNVVYNL
OC43	GGAVCLKHAE	EYREYLESYN	TATTAGFTFW	VYKTFDFYNL	WNTFTK---L	QSLNVVYNL
MHV	GGAVCLKHAE	DYREYLESYN	TATTAGFTFW	VYKTFDFYNL	WNTFTK---L	QSLNVVYNL
AIPV	GGAVCKKHAQ	MYAEFVTSYN	AAVTAGFTFW	VTNKLNPYNL	WKSFSF---L	QSIDNIAYNM
65 SARS CoV	GGAVCRHHAN	EYRQYLDAYN	MMISAGFSLW	IYKQFDTYNL	WNTFTK---L	QSLNVAYNV
	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	2105	2115	2125	2135	2145	2155
70 EMCR	VKKGCFGTVD	GELPVAIVND	KVFVRVGDVD	NLVFTNKTTL	PTNVAFELFA	KRKMGLTPPL
229E	VKKGCFVGD	GELPVAISGD	KVFVRDGTND	NLVFVNKTSI	PTNIAFELFA	KRKVGLTPPL
PEDV	LKKGCFVGD	GELPVAIVND	KVLVRDGTVD	TLVFTNKTSI	PTNVAFELFA	KRKVGLTPPI
TGEV	VKKGAFGTGLK	GDLPTAVIAD	KIMVRDGTVD	KCIFTNKTSI	PTNVAFELFA	KRKLGLTPPL
BoCoV	VKTGHYTGQA	GEMPCAIND	KVVAIDKED	VVIFINNTTY	PTNVAFELFA	KRSIRHHPPL
OC43	VKTGHYTGQA	GEMPCAIND	KVVAIDKED	VVIFINNTTY	PTNVAFELFA	KRSVRHHPPL
MHV	VNAGHFDGRA	GELPCAIVGE	KVIAIQONED	VVVFKNNTFF	PTNVAFELFA	KRSIRHHPPL
75 AIPV	VKGGHYDAIA	GEMPTVITGD	KVFVVDQVE	KAVFVNQTTL	PTSVAFELFA	KRNIRTPPNN
SARS CoV	VNKGHFDGHA	GEAPVSIINN	AVYTKVDGID	VEIFENKTTL	PVNVAFELFA	KRNIPVPEI
	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	2165	2175	2185	2195	2205	2215
80 EMCR	SILKNLGVVA	TYKFVLWDYE	AERFETSYSK	SVCKYTDEN-	-----EDV	CVCFDINSIQG



57/07

5	229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	SILKNLGVVA TILRNLGVVC TILRNLGVVA KLFRNLNIDV KLFRNLNIDV KLFRNLNIDV RILKGLGVDV KILNNLGVDI	TYKFVLWDYE TSKCVIWDYE TYKFVLWDYE CWKHVIWDYA CWKHVIWDYA CWSHVLWDYA TNGFVIWDYA AANTVIWDYK	AERPLTSFTK AERPLTTFETK AERPFNSFTK RESIFCSNTY RESIFCSNTY KDSVFCSSY NQTPLYRNTV REAPAHVSTI	SVCGYTDFA- DVCKYTDFE- QVCSYTDLD- GVCMTYDLK- GVCMTYDLK- KVCKYTDLQ- KVCAYTDIE- GVCTMTDIAR	-----EDV -----GDV -----SEV -----LIDKL -----FIDKL -----CIESL -----PNGL KPTESACSSL	CTCYDNSIQG CTLFDNSIVG VTCFDNSIAG NVLFDGRDNG NVLFDGRDNG NVLFDGRDNG VVLYDDR-YG TVLFDGRVEG
10		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		2225	2235	2245	2255	2265	2275
	EMCR 229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	SYERFTLTNN SYERFTLTSTN SLERFSMTQN SFERFTTTRD ALEAFKRSNN ALEAFKRSNN ALEAFKKCRD DYQSFLAADN QVDLFRNARN	AVLFSTVVIK AVLFSATAVK AVLMSLTAVK AVLISNNAVK GVYISTTKVK GVYISTTKVK GVYINTTKIK AVLVSTQCCK GVLITEGSVK	N-----LTPIK TGGKSLPAIK K-----LTGIK G-----LSAIK S-----LSMIR S-----LSMIR S-----LSMIK R-----YSYVE G-----LTPSK	LNFGMLNGMP LNFGMLNGNA LTGYGLNGVP LQYGLLNDLP GPPRAELNGV GPPRAELNGV GPQRADLNGV IPSNLLVQNG GPAQASVNGV	VSSIKSDKGV IATVKSEDGN VNTHED---- VSTVGN---- VVDKVGDD-- VVDKVGDD-- VVEKVGDD-- MPLKDG---- TLIGES----	EKLNVWYTYV IKNNINWVYV -KPTWYIYV -KPTWYIYV -TDCVFYFAV -TDCVFYFAV -SDVEFWFAM -ANLYVYK -VKTQFNHYK
20		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		2285	2295	2305	2315	2325	2335
	EMCR 229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	RKNG----- RKDG----- RKNG----- RKNG----- RKEGQDVIFS RKEGQDVIFS RRDGDVIFS RVNG----- KVDG-----	-----QFQDH -----KPVDD -----KFEDY -----EYVEQ QFDSLVRSSN QFDSLVRSSN RTGSLEPSHY -----AFVTL -----IIQQL	Y----- Y----- P----- I----- QSPQGNLGSN QSPQGNLGSN RSPQGNPGGN P----- P-----	----- ----- ----- ----- -EPGNVGGND GKPGNVGGND -RVGDLGNE ----- -----	-----DGFTYQ -----DGFTYQ -----DGFTYQ -----DSYTYQ ALATSTIFTQ ALATSTIFTQ ALARGTIFTQ -----NTINTQ -----ETYFTQ	GRNLSDFTPR GRNLQDFLPR GRTTADFSPR GRTFETFKPR SRVISSFTCR SRVISSFTCR SRFLSSFAPR GRSYETFEPR SRDLEDFKPR
25		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		2345	2355	2365	2375	2385	2395
	EMCR 229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	SDMEYDFLNM STMEEDFLNM SDMEKDFLSM STMEEDFLSM TDMEKDFIAL TDMEKDFIAL TDMEKDFIAL SEMEKDFMDL SDIERDFLAM SQMETDFLEL	DMGVFIQYK DIGVFIQYK DMGLFIQYK DTLFIQYK DQDVFIQYK DQDVFIQYK DQDVFIQYK DEDFVFIQYK SEESFVERYG AMDEFIQRK	LEDFNFEHV LEDFNFEHV LEDYGFHEVV LEDYGFHEVV LEDYAFEHIV LEDYAFEHIV LEDYAFEHIV LQDYAFEHVV -KDLGLQHIL LEGYAFEHIV	YGDVSKTTLG YGDVSKTTLG YGDVSKTTLG FGDVSKTTIG YGNFNQKIIG YGNFNQKIIG YGNFNQKIIG YGSFNQKIIG YGEVDKPKQLG YGDVSHGQLG	GLHLLISQFR GLHLLISQVR GLHLLISQVR GMHLLISQVR GLHLLIGLYR GLHLLIGLYR GLHLLIGLYR GLHLLIGLAR GLHTVIGMYR GLHLMIGLAR	LSKMGVLKAD LSKMGILKAE LACMGVLKID LAKMGLFSVQ RQQTNSNLVIQ RQQTNSNLVQV RQKNSNLVQV LLRANKLNK RSQDSPLKLE
30		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		2405	2415	2425	2435	2445	2455
	EMCR 229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	DEFVTASDTL EFVAASDTL EFVSSNDSTL EFMNSNDSTL EFVS-YDSSI EFVS-YDSSI EFVS-YDSSI EFVP-YDSSI SVTN-SDSDV DFIP-MDSTV	RCCTVTYLNE KCCTVTYLND KSCCTVYADN KSCCTVYADN HSYFITDEKS HSYFITDEKS HSYFITDEKS HSYFITDENS MONYFVLSN KNYFITDAQT	LSSKVVCTYM PSSKTVCTYM PSSKMVCTYM PSSKMVCTYM GGSKSVCTVI GGSKSVCTVI GGSKSVCTVI GGSKSVCTVI GSYKQVCTVV GSSKVCVSVI	DLLDDDFVTI DLLDDDFVSV DLLDDDFVSI DILLDDFVSI DILLDDFVAL DILLDDFVAL DILLDDFVAL DILLDDFVDI DILLDDFLEL DILLDDFVEI	LK---SLDLG LK---SLDIT LK---SLDLS IK---SLDGN VK---SLNLN VK---SLNLN VK---SLNLN VK---SLNLN LRNLIKVEYGT IK---SQDLS	VISKVHEVII VVKVHEVII VVKVHEVMV VVKVVDVIV CVSKVVNVNV CVSKVVNVNV CVSKVVNVNV CVSKVVNVNV NKSXVTVSVI VISKVVKVTI
35		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		2465	2475	2485	2495	2505	2515
	EMCR 229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	DNKPYRWMLW DNKPYRWMLW DCKMWRWMLW DCKMWRWMLW DFKDFQFMLW DFKDFQFMLW DFKDFQFMLW DYHSINFMFW DYABISEMLW	CKDNHLSTFY CKDNVATFY CKDHKLQTFY CENSHIKTFY CNDEKVMTFY CNDEKVMTFY CNDEKVMTFY FEDGSIKTCY CKDGHVETFY	PQLQS-AEWK PQLQS-AEWK PQLQA-SEWK PQLQA-AEWN PRLQAASDWK PRLQAASDWK PRLQAASDWK PQLQS-AWT PKLQASQAWQ	CGYAMPQIYK CGYSMPGIYK CGYSMPSIYK PGYSMPPTLYK PGYSMPVLYK PGYSMPVLYK PGYSMPVLYK CGYNMPELYK PGVAMPNLYK	LQRMCLPCN TQRMCLPCN IQRMCLPCN IQRMCLERCN YLNSEMERVS YLNSEMERVS YLNSEMERVS VQNCVMEPCN MQRMILLEKCD	LYNYGAGIKL LYNYGAGLKL LYNYGAGVKL LYNYGAGVKL LWNYGKPVTL LWNYGKPVTL LWNYGKPVTL IPNYGVGITL LQNYGENAVI
40		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		2525	2535	2545	2555	2565	2575
	EMCR 229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	PSGIMLNVVK PSGIMENVVK PDGIMENVVK PDGITNNVVK PTGCMNVAK PTGCMNVAK PTGCLMNVAK PSGIMLNVAK PKGIMNVAK	YTQLCQYLNS YTQLCQYFNS YTQLCQYLNS YTQLCQYLNT YTQLCQYLNT YTQLCQYLNT YTQLCQYLNT YTQLCQYLSK YTQLCQYLNT	TTMCVPHNMR TTLCVPHNMR TTMCVPHHMR TTLCVPHKMR TTTAVPVNTR TTTAVPVNTR TTTAVPANMR TTICVPHNMR LTLAVPVNMR	VLHYGAGSDK VLHLGAGSDY VLHLGAGSDK VLHLGAAGAS VLHLGAGSEK VLHLGAGSEK VLHLGAGSDK VMHFGAGSDK VIHFGAGSDK	GVAPGTTVLK GVAPGTAVLK GVAPGTAVLR GVAPGSTVLR GVAPGSAVLR GVAPGSAVLR DVAPGSAVLR GVAPGSTVLK GVAPGTAVLR	RWLPPD---- RWLPHD---- RWLPLD---- RWLPPD---- QWLPGATILR QWLPGATILR QWLPGATILR QWLPEG---- QWLPTG----
45		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		2525	2535	2545	2555	2565	2575
	EMCR 229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	PSGIMLNVVK PSGIMENVVK PDGIMENVVK PDGITNNVVK PTGCMNVAK PTGCMNVAK PTGCLMNVAK PSGIMLNVAK PKGIMNVAK	YTQLCQYLNS YTQLCQYFNS YTQLCQYLNS YTQLCQYLNT YTQLCQYLNT YTQLCQYLNT YTQLCQYLNT YTQLCQYLSK YTQLCQYLNT	TTMCVPHNMR TTLCVPHNMR TTMCVPHHMR TTLCVPHKMR TTTAVPVNTR TTTAVPVNTR TTTAVPANMR TTICVPHNMR LTLAVPVNMR	VLHYGAGSDK VLHLGAGSDY VLHLGAGSDK VLHLGAAGAS VLHLGAGSEK VLHLGAGSEK VLHLGAGSDK VMHFGAGSDK VIHFGAGSDK	GVAPGTTVLK GVAPGTAVLK GVAPGTAVLR GVAPGSTVLR GVAPGSAVLR GVAPGSAVLR DVAPGSAVLR GVAPGSTVLK GVAPGTAVLR	RWLPPD---- RWLPHD---- RWLPLD---- RWLPPD---- QWLPGATILR QWLPGATILR QWLPGATILR QWLPEG---- QWLPTG----
50		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		2525	2535	2545	2555	2565	2575
	EMCR 229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	PSGIMLNVVK PSGIMENVVK PDGIMENVVK PDGITNNVVK PTGCMNVAK PTGCMNVAK PTGCLMNVAK PSGIMLNVAK PKGIMNVAK	YTQLCQYLNS YTQLCQYFNS YTQLCQYLNS YTQLCQYLNT YTQLCQYLNT YTQLCQYLNT YTQLCQYLNT YTQLCQYLSK YTQLCQYLNT	TTMCVPHNMR TTLCVPHNMR TTMCVPHHMR TTLCVPHKMR TTTAVPVNTR TTTAVPVNTR TTTAVPANMR TTICVPHNMR LTLAVPVNMR	VLHYGAGSDK VLHLGAGSDY VLHLGAGSDK VLHLGAAGAS VLHLGAGSEK VLHLGAGSEK VLHLGAGSDK VMHFGAGSDK VIHFGAGSDK	GVAPGTTVLK GVAPGTAVLK GVAPGTAVLR GVAPGSTVLR GVAPGSAVLR GVAPGSAVLR DVAPGSAVLR GVAPGSTVLK GVAPGTAVLR	RWLPPD---- RWLPHD---- RWLPLD---- RWLPPD---- QWLPGATILR QWLPGATILR QWLPGATILR QWLPEG---- QWLPTG----
55		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		2525	2535	2545	2555	2565	2575
	EMCR 229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	PSGIMLNVVK PSGIMENVVK PDGIMENVVK PDGITNNVVK PTGCMNVAK PTGCMNVAK PTGCLMNVAK PSGIMLNVAK PKGIMNVAK	YTQLCQYLNS YTQLCQYFNS YTQLCQYLNS YTQLCQYLNT YTQLCQYLNT YTQLCQYLNT YTQLCQYLNT YTQLCQYLSK YTQLCQYLNT	TTMCVPHNMR TTLCVPHNMR TTMCVPHHMR TTLCVPHKMR TTTAVPVNTR TTTAVPVNTR TTTAVPANMR TTICVPHNMR LTLAVPVNMR	VLHYGAGSDK VLHLGAGSDY VLHLGAGSDK VLHLGAAGAS VLHLGAGSEK VLHLGAGSEK VLHLGAGSDK VMHFGAGSDK VIHFGAGSDK	GVAPGTTVLK GVAPGTAVLK GVAPGTAVLR GVAPGSTVLR GVAPGSAVLR GVAPGSAVLR DVAPGSAVLR GVAPGSTVLK GVAPGTAVLR	RWLPPD---- RWLPHD---- RWLPLD---- RWLPPD---- QWLPGATILR QWLPGATILR QWLPGATILR QWLPEG---- QWLPTG----
60		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		2525	2535	2545	2555	2565	2575
	EMCR 229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	PSGIMLNVVK PSGIMENVVK PDGIMENVVK PDGITNNVVK PTGCMNVAK PTGCMNVAK PTGCLMNVAK PSGIMLNVAK PKGIMNVAK	YTQLCQYLNS YTQLCQYFNS YTQLCQYLNS YTQLCQYLNT YTQLCQYLNT YTQLCQYLNT YTQLCQYLNT YTQLCQYLSK YTQLCQYLNT	TTMCVPHNMR TTLCVPHNMR TTMCVPHHMR TTLCVPHKMR TTTAVPVNTR TTTAVPVNTR TTTAVPANMR TTICVPHNMR LTLAVPVNMR	VLHYGAGSDK VLHLGAGSDY VLHLGAGSDK VLHLGAAGAS VLHLGAGSEK VLHLGAGSEK VLHLGAGSDK VMHFGAGSDK VIHFGAGSDK	GVAPGTTVLK GVAPGTAVLK GVAPGTAVLR GVAPGSTVLR GVAPGSAVLR GVAPGSAVLR DVAPGSAVLR GVAPGSTVLK GVAPGTAVLR	RWLPPD---- RWLPHD---- RWLPLD---- RWLPPD---- QWLPGATILR QWLPGATILR QWLPGATILR QWLPEG---- QWLPTG----
65		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		2525	2535	2545	2555	2565	2575
	EMCR 229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	PSGIMLNVVK PSGIMENVVK PDGIMENVVK PDGITNNVVK PTGCMNVAK PTGCMNVAK PTGCLMNVAK PSGIMLNVAK PKGIMNVAK	YTQLCQYLNS YTQLCQYFNS YTQLCQYLNS YTQLCQYLNT YTQLCQYLNT YTQLCQYLNT YTQLCQYLNT YTQLCQYLSK YTQLCQYLNT	TTMCVPHNMR TTLCVPHNMR TTMCVPHHMR TTLCVPHKMR TTTAVPVNTR TTTAVPVNTR TTTAVPANMR TTICVPHNMR LTLAVPVNMR	VLHYGAGSDK VLHLGAGSDY VLHLGAGSDK VLHLGAAGAS VLHLGAGSEK VLHLGAGSEK VLHLGAGSDK VMHFGAGSDK VIHFGAGSDK	GVAPGTTVLK GVAPGTAVLK GVAPGTAVLR GVAPGSTVLR GVAPGSAVLR GVAPGSAVLR DVAPGSAVLR GVAPGSTVLK GVAPGTAVLR	RWLPPD---- RWLPHD---- RWLPLD---- RWLPPD---- QWLPGATILR QWLPGATILR QWLPGATILR QWLPEG---- QWLPTG----
70		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		2525	2535	2545	2555	2565	2575
	EMCR 229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	PSGIMLNVVK PSGIMENVVK PDGIMENVVK PDGITNNVVK PTGCMNVAK PTGCMNVAK PTGCLMNVAK PSGIMLNVAK PKGIMNVAK	YTQLCQYLNS YTQLCQYFNS YTQLCQYLNS YTQLCQYLNT YTQLCQYLNT YTQLCQYLNT YTQLCQYLNT YTQLCQYLSK YTQLCQYLNT	TTMCVPHNMR TTLCVPHNMR TTMCVPHHMR TTLCVPHKMR TTTAVPVNTR TTTAVPVNTR TTTAVPANMR TTICVPHNMR LTLAVPVNMR	VLHYGAGSDK VLHLGAGSDY VLHLGAGSDK VLHLGAAGAS VLHLGAGSEK VLHLGAGSEK VLHLGAGSDK VMHFGAGSDK VIHFGAGSDK	GVAPGTTVLK GVAPGTAVLK GVAPGTAVLR GVAPGSTVLR GVAPGSAVLR GVAPGSAVLR DVAPGSAVLR GVAPGSTVLK GVAPGTAVLR	RWLPPD---- RWLPHD---- RWLPLD---- RWLPPD---- QWLPGATILR QWLPGATILR QWLPGATILR QWLPEG---- QWLPTG----
75		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		2525	2535	2545	2555	2565	2575
	EMCR 229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	PSGIMLNVVK PSGIMENVVK PDGIMENVVK PDGITNNVVK PTGCMNVAK PTGCMNVAK PTGCLMNVAK PSGIMLNVAK PKGIMNVAK	YTQLCQYLNS YTQLCQYFNS YTQLCQYLNS YTQLCQYLNT YTQLCQYLNT YTQLCQYLNT YTQLCQYLNT YTQLCQYLSK YTQLCQYLNT	TTMCVPHNMR TTLCVPHNMR TTMCVPHHMR TTLCVPHKMR TTTAVPVNTR TTTAVPVNTR TTTAVPANMR TTICVPHNMR LTLAVPVNMR	VLHYGAGSDK VLHLGAGSDY VLHLGAGSDK VLHLGAAGAS VLHLGAGSEK VLHLGAGSEK VLHLGAGSDK VMHFGAGSDK VIHFGAGSDK	GVAPGTTVLK GVAPGTAVLK GVAPGTAVLR GVAPGSTVLR GVAPGSAVLR GVAPGSAVLR DVAPGSAVLR GVAPGSTVLK GVAPGTAVLR	RWLPPD---- RWLPHD---- RWLPLD---- RWLPPD---- QWLPGATILR QWLPGATILR QWLPGATILR QWLPEG---- QWLPTG----
80		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		2525	2535	2545	2555	2565	2575
	EMCR 229E PEDV TGEV BoCoV OC43 MHV AIPV SARS CoV	PSGIMLNVVK PSGIMENVVK PDGIMENVVK PDGITNNVVK PTGCMNVAK PTGCMNVAK PTGCLMNVAK PSGIMLNVAK PKGIMNVAK	YTQLCQYLNS YTQLCQYFNS YTQLCQYLNS YTQLCQYLNT YTQLCQYLNT YTQLCQYLNT YTQLCQYLNT YTQLCQYLSK YTQLCQYLNT	TTMCVPHNMR TTLCVPHNMR TTMCVPHHMR TTLCVPHKMR TTTAVPVNTR TTTAVPVNTR TTTAVPANMR TTICVPHNMR LTLAVPVNMR	VLHYGAGSDK VLHLGAGSDY VLHLGAGSDK VLHLGAAGAS VLHLGAGSEK VLHLGAGSEK VLHLGAGSDK VMHFGAGSDK VIHFGAGSDK	GVAPGTTVLK GVAPGTAVLK GVAPGTAVLR GVAPGSTVLR GVAPGSAVLR GVAPGSAVLR DVAPGSAVLR GVAPGSTVLK GVAPGTAVLR	RWLPPD---- RWLPHD---- RWLPLD---- RWLPPD---- QWLPGATILR QWLPGATILR QWLPGATILR QWLPEG---- QWLPTG----

58/87

		2585	2595	2605	2615	2625	2635
5	EMCR	-----AIII	DNDINDYVSD	ADFSITGDCA	TVYLEDKFDL	LISDMYDG--	--RIKFCDGE
	229E	-----AIVV	DNDVVDYVSD	ADFSVTGDCA	TVYLEDKFDL	LISDMYDG--	--RTKAIDGE
	PEDV	-----AIIV	DNDSDYVSD	ADYSVTGDCS	TLYLSDKFDL	VISDMYDG--	--KIKSCDGE
	TGEV	-----AILV	DNDLRDYVSD	ADFSVTGDCT	SLYIEDKFDL	LVSDLYDG--	--STKSIDGE
	BoCoV	QWLPA GTILV	HNDLYPFVSD	SVATYFGDCI	TLFPDCQWDL	IISDMYD---	--LLLDIGVH
	OC43	-----TILV	DNDLYPFVSD	SVATYFGDCI	TLFPDCQWDL	IISDMYD---	--LTKNIGEV
10	MHV	-----SILV	DNDINPFVSD	SVASYGNCI	TLPIACQWDL	IISDMYD---	--LTKNIGEV
	AIPV	-----TLLV	DNDIVDYVSD	AHVSVLSDCN	KYNTCHKFDL	VISDMYTDND	SKRKHEGVIA
	SARS CoV	-----TLLV	DSDLNDFVSD	ADSTLIGDCA	TVHTANKWDL	IISDMYD---	--RTKHVTKE
15	EMCR	2645	2655	2665	2675	2685	2695
	229E	NVSKDGFFTY	LNGVIREKLA	IGGSVAIKIT	EYSWNKLYE	LIQRFQFRTL	FCTSVNTSSS
	PEDV	NVSKEGFFTY	INGFICEKLA	IGGSTAIKVT	EYSWNKKLYE	LVQRFSEFTW	FCTSVNTSSS
	TGEV	NVSKEGFFTY	INGVITEKLA	LGSTVAIKVT	EFSWNKKLYE	LIQRFQFRTL	FCTSVNTSSS
	BoCoV	NTSKDGFFTY	INGFIKEKLS	LGGSSVAIKIT	EFSWNKDLYE	LIQRFQFRTL	FCTSVNTSSS
20	OC43	VVRCS---YI	HCHMIRDKLA	LGGSSVAIKIT	EFSWNAELYK	LMGYFAFWTV	FCTNANASSS
	MHV	NVSKDGFFTY	ICHMIRDKLA	LGGSSVAIKIT	EFSWNAELYK	LMGYFAFWTV	FCTNANASSS
	AIPV	NVSKDGFFTY	LCHLIRDKLA	LGGSSVAIKIT	EFSWNAELYS	LMGKFAFWTV	FCTNANASSS
	SARS CoV	NNGNDVFIY	LSSFLRNLA	LGGSSFAVKVT	ETSWHEVLYD	IAQDCAWWTM	FCTAVNASSS
25		NDSKEGFFTY	LCGFIKQKLA	LGGSSIAVKIT	EHSWNADLYK	LMGHFSWWTM	FVTNANASSS
30	EMCR	2705	2715	2725	2735	2745	2755
	229E	EAFLLIGINYL	GDFIQGFFIA	GNTVHANYIF	WRNSTIMSLS	YNSVLDLSKF	ECKHKATVVV
	PEDV	EAFVVGINYL	GDFIAQGFID	GNITHANYVF	WRNSTVMSLS	YNSVLDLSKF	NCKHKATVVV
	TGEV	EAFLLIGVHYL	GDFASGAVID	GNTMHANYIF	WRNSTIMTMS	YNSVLDLSKF	NCKHKATVVV
	BoCoV	EGFLIGINYL	GPYCDKAIIV	GNIMHANYIF	WRNSTIMALS	YNSVLDTPKF	KCRCNALIV
	OC43	EGFLIGINYL	GK--PKVEID	GNVMHAIICF	G-----	-----	EIPQFGTGV
	MHV	EGFLIGINYL	CK--PKVEID	GNVMHANYLF	WRNSTVWNGG	AYSLFDMAKF	PLKLAGTAVI
35	AIPV	EGFLIGINYL	NR--TRTEID	GNVMHANYLF	WRNSTVWNGG	AYSLFDMSKF	PLKLAGTAVI
	SARS CoV	EAFLLIGVNYL	GAS-EKVKVS	GKTLHANYIF	WRNCNYLQTS	AYSIFDVAKF	DLRLKATPVV
40		EAFLLIGANYL	GK--PKEQID	GYTMHANYIF	WRNTNPIQLS	YSYSLFDMSKF	PLKLRGTAVM
45	EMCR	2765	2775	2785	2795		
	229E	TLKDSVDNDM	VLSLIKSGRL	LLRNSGRFGG	FSNHLVSTK-		
	PEDV	QLKDSIDINEM	VLSLVRSGKL	LVRGNGKCLS	FSNHLVSTK-		
	TGEV	NLKDSSISDV	VLGLLKNGKL	LVRNDAICG	FSNHLVNVNK		
	BoCoV	NLKEKELNEM	VIGLLRKGL	LIRNNGKLLN	FGNHVNTF-		
	OC43	IACLIWLNSR	LSWLMP---	-----	-----		
	MHV	NLRADQINDM	VYSLEKGGK	LIRDINKEVF	VGDSLNVNI-		
	AIPV	SLKPDQINDL	VLSLIEKGGK	LVRDTRKEVF	VGDSLNVNK-		
	SARS CoV	NLKTEQKTDL	VFNLIKCGKL	LVRDVGNTSF	TSDFVCTM-		
50		SLKENQINDM	IYSLEKGGRL	IIRENNRVVV	SSDILVNN--		

#### d. Putative Orf lab

		5	15	25	35	45	55
55	EMCR	-----	-----M	FYNQVTLAVA	SDSEISGFGF	AIPSVAVRAY	SEAAAQGFQA
	229E	-----	-----M	ACNRVTLAVA	SDSEISANGC	STIAQAVRRY	SEAAASNGFRA
	PEDV	-----	-----M	ASNHVTLAVA	NDAEISAFGF	CTASEAVSY	SEAAASGFMQ
	TGEV	-----	-----M	SSKQFKILVN	EDYQVNVPSL	PIR-DVLQEI	KYCYRNGFEG
60	OV43	MSKINKYGLE	LHWAPEFPWM	FEDAEEKLDN	PSSSEVDMIC	STTAQKLETG	GICPENHVMV
	BoCoV	MSKINKYGLE	LHWAPEFPWM	FEDAEEKLDN	PSSSEVDIVC	STTAQKLETG	GICPENHVMV
	MHV	MAKMGKYGLG	FKWAPEFPWM	LPNASEKLGS	PERSEEDGFC	PSAAQEPKTK	GKTLINHVRV
	AIBV	-----	-----	MASSLKQGV	PKPRDVILVS	KDIPQLCDA	LFYFYTSHNPK
	SARS CoV	-----	--MESLVIGV	NEKTHVQLSL	PVLQVRDVLV	RGFGDSVEEA	LSEAREHLKN
65	EMCR	65	75	85	95	105	115
	229E	CRFVAFGLQD	CVTGINDDDY	VIALTG---	-----	TNQLCAKILL	FSDRPLNLRG
	PEDV	CRFVSLDLQD	CIVGIADDTY	VMGLHG---	-----	NQTLFCNIMK	FSDRPFMLHG
	TGEV	CRFVSLDLAD	TVEGLLPEDY	VMVVIG---	-----	TTKLSAYVDT	FGSRPRNICG
	OV43	YVFVPEYCRD	LVDCKDKHY	VIGVLG---	-----	NGVSDLKPV	LTEPSVMLQG
	BoCoV	DCRRLKQEC	CVQSSILIREI	VMNASPYDLE	VLLQDALQSR	EAVLVTPPLG	MSLEACYVRG
	MHV	DCRRLKQEC	CVQSSILIREI	VMNTRPYDLE	VLLQDALQSR	EAVLVTPPLG	MSLEACYVRG
	AIBV	DCSRLPALEC	CVQSAIIRDI	FVDEDPNVE	ASTMMALQFG	SAVLVKPSKR	LSIQAWAKLG
75	SARS CoV	DYADAFAVRQ	KFDRSLQTKG	QFKFET---	-----V	CGLFLKQVD	KITPGVPAKV
		GTCGLVELEK	GVLPLQEQPY	VFIKR--SDA	LSTNHGHKVV	ELVAEMDGIQ	YGRSGITLGV
		125	135	145	155	165	175

59/87

5	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	WLIFSNSNYV LQDFDVVFG--HGAGSVVVFV DKYMCQFDDGK PVLP--KNMW EFRDYFNDNT WLVFSNSNYL LEEFDVVFGK -RGGGNVTYT DQYLCGADGK PVMS--EDLW QFVDHFGENE WLLFSNCNYF LEELELTFG--RRGGNIVPV DQYMCQADGK PVLQ--ESEW EYTDFFADSE FIVRANCNGV LEDFDLKI--RTGRGAIYV DQYMCQADGK PVIE--G--- DFKDYFGDED CNPKGWTMGL FRRRSVCNTG RCTVKNKHVAY QLYMIDPAGV CLGAGQFVGW VIPLAFMPVQ CNPNGWTMGL FRRRSVCNTG RCAVNKHVAY QLYMIDPAGV CFCGAGQFVGW VIPLAFMPVQ VLPKTPAMGL FKRFCLCNTR ECVCDAHVAF QLFVTPQPDGV CLNGRQFVGW FVPVTAIPAY LKATSKLADL EDIFGVSPLA RKYRELLKTA CQWLTVEAL DVV-----AQ TLDEIFDPTPE LVPHVGETPI AYRNVLLRK-- --NGNKGAGG HSYGIDLKSY DLG--DELGT DPIDYEQNWN
10		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....  185 195 205 215 225 235
15	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	DS-IVIGGVT YQLAWDVIRK DLSYEQQNVL AIESIHYLG- TTGHTLKS GC KLINAKPPKY E--IIINGHT YVCAWLTKRK PLDYKRONNL AIEIEYVHG DALHTLRNGS VLEMAKEVKT DQGLNIAGIT YVCAWIVERS DVSASQNL SIKSITYCS- TYEHTFLDGT AMKVARTPKI --IIEFEGEE YHCAWTTVRD EKPLNQQTLE TIQEIQYNL- DIPHKLPNCA TRHVAPPVKK SRKFIVPWVM YLRKRGEKGA YNKDHGRGGE GH-VYDFKVE DAYDQVHDEP KGKFSKKAYA SRKFIVPWVM YLRKRGEKGA YIKDYKRGGF EH-VYNFKVE DAYDLVHDEP KGKFSKKAYA AKQWLQWPSI LLRKGGNKGVS VTSGHFRRRAV TMPVYDFNVE DACEVHLNP KGKYSRKAYA ILWLQVAAKI HVSSMAMRRL VGEVTAQVMD ALG----- SNLSALFQIV KQIARIFQK NTKHGSGALR ELTRELNNGA VTRYVDNNFC GPDGYPLDCI KDFLARAGKS MCTLS-EQLD
20		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....  245 255 265 275 285 295
25	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	SSKVVLSEGE NAVYKAFGSP FITNGISLLD IIVKPVFFNA FVKCNCGSEN WSVGAWDGYL SSKVVLSDAL DKLYKVFSGP VMTNGSNILE AFTKPVFISA LVQCTCKGTS WSVGDWTGFK KKNVVLSEPL ATIYREIGSP FVNGSDARS IIRRPVFLHA FVKCKCGSYH WTVGDWTSYV NSKIVLSEDY KKLYDIFGSP FMGNGDCLSK CFTDLHFIAA TLRCPCGSES SGVGDWTGFK LIRGYRGVGP LLYVDQYQCD YTGSLADGLE AYADKTLOEM KALFPTWSQE LLFDVIVAWH LIRGYRGVGP LLYVDQYQCD YTGSLADGLE AYADKTLOEM KALFPTWSQE LPFDVTVAWH LLKGYRGVKS ILFLDQYQCD YTGSLADGLE AYADKTLOEM KALFPTWSQE LDNEVVVAWH ALAFIFENVNE LPQRIALKM AFACARISIT VVVVERTLVV KEFAGTCLAS INGAVAKFFE YIESKRGVYC CRDHEHEIAW FTERSDKSYE HQTFFEIFKSA KKEDTFKGEK PKFVFPNLSK
30		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....  305 315 325 335 345 355
35	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	SSCCGTFPAKK LCVVPGNVVP GDVITSTDA GCGVKYAGL VVKHITNITG VSLWRVTAVH SSCCNVISNK LCVVPGNVKP GDAVITQQA GAGIKYFCGM TLKFVANIEG VSVWRVIALQ STCCGFKCKP VLVASCAMP GSVVVTTRAGA GTGVKYNNM FLRHVADIDG LAFWRILKQ TACCGLSGKV KGVTLGDIKP GDAVVTMSA GKGVKFFANC VLQYAGDVGE VSIWKVITF VVRDPRVVMR LQSAATIRSV AYVANPTEDL CDGSVVIKEP VHVYADDSII LRQYNLVDIM VVRDPRVVMR LQSAATIRSV AYVANPTEDL CDGSVVIKEP VHVYADDSII LRQYNLVDIM VVRDPRVVMR LQSAATIRSV AYVANPTEDL CDGSVVIKEP VHVYADDSII LRQYNLVDIM VVRDPRVVMR LQSAATIRSV AYVANPTEDL CDGSVVIKEP VHVYADDSII LRQYNLVDIM ELPENGFMGSK IFTTLAFFKE AAVRVVENIP VDGDDVVREP AHLAANAIV KRLPRLVETM VKVIQPRVEK KKTGEGFMGRI RSVYVPVASPO ECNNMHLSTL MKCNHCDEVS WQTCDFLKAT
40		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....  365 375 385 395 405 415
45	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	SDGMFVATSS YDALLHRNSL DFFCFDVNTL LSNQLRLAFL GASVTEDVKF A-----AST SVDCFEVASS FVEEHEVNRM DTFCFVNRNS VTDECLAML GAEMTSNVR Q-----VAS SKDDLACSGK FLEHHEEGFT DPCYFLNDSS LATKLKFDIL SGKFSDEVKQ A-----IA TVDETVCPTG FEDELN---- DFIKPEKSL VACSVKRAFI TGDIDDAVDH C-----IIT SHFYMEADTV VNAFYGVALK DCGFVMQFGY IDCEQDSCDF KGWIPGNMID G-----FACTTC SCFYMEADAV VNAFYGVLDL DCGFVMQFGY IDCEQDSCDF KGWIPGNMID G-----FACTTC LYT-----DSSV TEFCEYTKLC DCGFTTQFGY VDCCGDACDF RGWVPGNMMD G-----FLCPGC LLDQKADIPV EPEGWS---- --AILDGHLC YVFRSGDRFY AAPLSGNFAL S----- CEHCGTENLV IEGPTTCGYL PTNAVVKMPC PACQDPEIGP EHSVADYHNN SNIETRLRKG
50		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....  425 435 445 455 465 475
55	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	GVIDISAGMF GLYDDILTNN KPWFVRKASG LFDATWDAFV AAIKLVPTTT GGLVRFVKSI GVIDISTGWF DVIDDIFAES KPWFVRKASG LFDATWDAFV AAIKLVPTTT GGLVRFVKSI GHVVVGSALV DIVDDALG-- QPWFIRKLGD IFGPCWSALA SALKQLKVTT GGLVRFVKSI GKLDLSTNLF GNVGLLFKK- TPWFVQKCGA LFDATWDAFV AAIKLVPTTT GGLVRFVKSI GHVYETGDLI AQSSGVLPVN PVLHTKSAAG YGGFGCKDSF TLYGQTVVYF GGCYVWSPAR GHVYETGDLI AQSSGVLPVN PVLHTKSAAG YGGFGCKDSF TLYGQTVVYF GGCYVWSPAR SKSYMPWELE AQSSGVLPVN PVLHTKSAAG YGGFGCKDSF TLYGQTVVYF GGCYVWSPAR -DVHCCERVV CLSDGVTP-- --EIN--DGL ILAAYSSFS VSELVTALKK GEPFKFLGHK GRTRCFGGCV FAYVGCYNKR AYWVPRASAD IG----SGHT GITGDNVETL NEDLLEILSR
60		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....  485 495 505 515 525 535
65	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	ASTVLTVSNG VIIMCADVPD AFQPVYRTFT QAICAAFDPS LDVFKIG-- CNSAVAVVGG TIQILASVPE KFLNAFDVVF TAIQTVFDCA VETCTIA-- SCATLSIVNG VEEFLADVPE KLAATAVTFV NFLNEFFESA CDCCLKVG-- CTSAFTIVNY KPTFVVPD-N RVKDLVOKCV KVLVKAQDV TQITITAG-- NIWIPILKSS VKSYDSLVT GVLGCKAIVK ETNLICKALY LDYVQHKCGN LHQRELLGVS NIWIPILKSS VKSYDSLVT GVLGCKAIVK ETNLICKALY LDYVQHKCGN LHQRELLGVS GMWLPVWISS VKSYADLYT GVVGCKAIVQ ETDAICRSY MDYVQHKCGN LEQRAILGLD ---FVYAKDA AVSFTLAKAA TIADVLRFLQ SARVIAEDVW SSFTEKS--
70		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....  485 495 505 515 525 535
75	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV	ASTVLTVSNG VIIMCADVPD AFQPVYRTFT QAICAAFDPS LDVFKIG-- CNSAVAVVGG TIQILASVPE KFLNAFDVVF TAIQTVFDCA VETCTIA-- SCATLSIVNG VEEFLADVPE KLAATAVTFV NFLNEFFESA CDCCLKVG-- CTSAFTIVNY KPTFVVPD-N RVKDLVOKCV KVLVKAQDV TQITITAG-- NIWIPILKSS VKSYDSLVT GVLGCKAIVK ETNLICKALY LDYVQHKCGN LHQRELLGVS NIWIPILKSS VKSYDSLVT GVLGCKAIVK ETNLICKALY LDYVQHKCGN LHQRELLGVS GMWLPVWISS VKSYADLYT GVVGCKAIVQ ETDAICRSY MDYVQHKCGN LEQRAILGLD ---FVYAKDA AVSFTLAKAA TIADVLRFLQ SARVIAEDVW SSFTEKS--
80		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....  485 495 505 515 525 535
	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV	ASTVLTVSNG VIIMCADVPD AFQPVYRTFT QAICAAFDPS LDVFKIG-- CNSAVAVVGG TIQILASVPE KFLNAFDVVF TAIQTVFDCA VETCTIA-- SCATLSIVNG VEEFLADVPE KLAATAVTFV NFLNEFFESA CDCCLKVG-- CTSAFTIVNY KPTFVVPD-N RVKDLVOKCV KVLVKAQDV TQITITAG-- NIWIPILKSS VKSYDSLVT GVLGCKAIVK ETNLICKALY LDYVQHKCGN LHQRELLGVS NIWIPILKSS VKSYDSLVT GVLGCKAIVK ETNLICKALY LDYVQHKCGN LHQRELLGVS GMWLPVWISS VKSYADLYT GVVGCKAIVQ ETDAICRSY MDYVQHKCGN LEQRAILGLD ---FVYAKDA AVSFTLAKAA TIADVLRFLQ SARVIAEDVW SSFTEKS--

60/87

SARS CoV		---ERVNINI VGDFHLNEEV AIIILASFSAS TSAFIDTIKS LDYKSFKT-I VESCGNYKVT					
		545	555	565	575	585	595
5	EMCR	---DVKFKR	LG DYVL TENA	LVRLTTEVVR	GVRD----	-----	---A----
	229E	---GKAFDK	VFDYVLLDNA	LVKLVTTKLK	GVRE----	-----	---R----
	PEDV	---GKTFNK	VGSYVLFDNA	LVKLVKAKAR	GPRQ----	-----	---A----
	TGEV	--IEAKCFVL	GAKYLLFNNA	LVKLVSVKIL	GKKQ----	-----	---K----
	OV43	DVWHKQLLLN	RGVYKPLEEN	IDYFNMRRAK	FSLETFT--	VCADGFMPFL	DDLVPRAYY
10	BoCoV	DVWHKQLLLN	RGVYKPLEEN	IDYFNMRRAK	FSLETFT--	VCADGFMPFL	DDLVPRAYY
	MHV	DVYHRQLLVN	RGDYSLLLEN	VDLFVKRRAE	FACK-FA---	TCGDGLVPLL	LDGLVPRSY
	AIBV	-----	FEFWKLAYGK	VRNLEEFVKT	YVCK-----	-----	-----
	SARS CoV	KGKPVKGAWN	IGQQRSVLTP	LCGFPSQAAG	VIRSIFARTL	DAANHSIPDL	QRAAVTILDG
15		605	615	625	635	645	655
	EMCR	-RIKKAMFTK	VVVGPTTEVK	FSVIELATVN	LRLVDCAPVV	CPKKGKIVVIA	GQAFFYSGGF
	229E	-GLNKVKYAT	VVVGSTEEVK	SSRVERSTAV	LTIANNSYKL	FDEGYTVVIG	DVAYFVSDGY
	PEDV	-GICEVRYTS	LVVGSTTKVV	SKRVENANVN	LVVVDEDTV	NTTGRTVVVD	GLAFFESDGF
20	TGEV	-GLECAFFAT	SLVGATVNVT	PKRTETATIS	LNKVDVAVP	G-EGYIVIVG	DMAFYKSGEY
	OV43	LAVSGQAFCD	YADKLCHAVV	SKSKEKLDVS	LDLSGAAIHY	LNSKIVDLAQ	HFSDFGTSEV
	BoCoV	LAVSGQAFCD	YAGKICHAVV	SKSKEKLDVS	VDSLGAIIHY	LNSKIVDLAQ	HFSDFGTSEV
	MHV	LIKSGQAFTS	MMVNFSHEVT	DMCMDMALLE	MHDVKVATKY	VKKVTGKLAV	RFKALGVAVV
	AIBV	AQMSIVILAA	VLGEDIWHLV	SQVIYKLGVL	FTKVVDFCDK	HWKGFVCVQLK	RAKLIVTETF
25	SARS CoV	ISEQSLRLVD	AMVYTSDLLT	NSVIIIMAYVT	GGLVQQTSSQ	LSNLLGTTFE	KLREIFEWIE
		665	675	685	695	705	715
	EMCR	YRFMVDSTTV	LNDFPVFTGEL	FYTIKFSGFK	LDGFN-----	---HQFVNAS	SATDAIIAVE
	229E	FRLMASPNVS	LTAVYKELF	AFNVNVMGTR	PE-----	---KFPTTV	TCENLESAVL
	PEDV	YRHLADADV	IEHPVYKSAC	ELKPVFECDD	IP--D-----	---FPLPVAA	SVABLVCQTD
	TGEV	YFMMSSPNFV	LTNNVFKAVK	VPSYDIVDV	DNDTKSKMIA	KLGSSEFYDG	DIDAAIVKVN
	OV43	SKIYHFFKTF	TTSTALAFW	VLPHVLHGAY	IVVESDIYFV	KN-IPRYASA	VAQAFQSVAK
	BoCoV	SKIYHFFKTF	TTSTALAFW	VLPHVLHGAY	IVVESDIYFV	KN-IPRYASA	VAQAFQSVAK
35	MHV	RKITEWFDLA	VDTAASAAGW	LCYQLVNGLE	AVANGGITFL	SD-VPELVKN	FVDKEKVFVK
	AIBV	CVLKGVQHC	FQLLLDAIHS	LYKSFKKCAL	GRIHG-----	---DLFWKGG	VHKIVQDGE
	SARS CoV	AKLSAGVEFL	KD-----	AWE	ILKFLITGVF	DIVKGQIQVA	SDNIKDCVKC
		725	735	745	755	765	775
40	EMCR	LLLSDFKTAV	FVYTCVVDGC	SVIVRRDAT-	FATHVCFKDC	YSIWEQFCID	NCGE-----
	229E	FLVNDKITEFQ	LDYSIDVIDN	EIIKPNIS-	LCVPLYVRDY	VDKWDDECRQ	YSNE-----
	PEDV	LLKKNYNTPY	KTYSCVVRGD	KCCITCTLO-	FKAPSYVEDA	VN-FVDLCTK	NIGT-----
	TGEV	ELLIEFRQOS	LCFRAFKDDK	SIFVEAYFKK	YKMPACLAH	IG-LWNIKK	DSCK-----
45	OV43	VVLDSLRTVF	IDGLSCFKIG	RRRICLSGRK	IYEVERG-LL	HSSQLPLDVY	DLTMPQSVQK
	BoCoV	VGLDSLRTVF	IDGLSCFKIG	RRRICLSGRK	IYEVERG-LL	HSSQLPLDVY	DLTMPQSVQK
	MHV	VLIDSMSVS	LSGLTVVKT	SNRVCLAGCK	VYEVVQK-RL	SAYVMPVGCN	EATC-----
	AIBV	IFWDAIDSVD	VEDLGVVQEK	SIDFEVCDV	TLPENQPGHM	VQIEDDGKNY	MFFR-----
	SARS CoV	MCIDQVTIAG	AK-LRSLNLG	EVFIAQSKGL	YRQCINGKEQ	LQLLMLPKAP	KEVT-----
50		785	795	805	815	825	835
	EMCR	-----PW	FLTDYNAILQ	SNNPQCAIVQ	ASESK-----	---VLLERFLP	KCPEILLSID
	229E	-----SW	FEDDYRAFIS	VLDITDAAVK	AAESK-----	---AFVDITVP	PCPSILKVID
	PEDV	-----AG	FHEFYITAHE	QQDLQGLT	CCTMSG----	F-ECFMPTIP	QCPAVLEEID
	TGEV	-----RG	FLNLFNHLNE	LEDIKETNIQ	AIKN-----	-----I-	LCPDPLDL
	OV43	AKQKPIYLKG	SGSDFSLADS	VVEVVTSLT	PCG-----YS	EPPKVADKIC	IVDNVYMAKA
	BoCoV	TKQKGIYILKG	SGSDFSLADS	VVEVVTSLT	PCG-----YS	EPPKVADKIC	IVDNVYMAKA
	MHV	-----LVG	ETEPAVVEDD	VVDVVKAPLT	YQG-----CC	KPPTSFEKIC	VVDKLYMAKC
60	AIBV	-----FKK	DENIYYTPMS	QLGAINVCK	AGG-----	-----KTVT	FGETTVQETP
	SARS CoV	-----FLEG	DSHDTVLTSE	EVVLKNGELE	ALETVPDSFT	NGAIVGTPVC	VNGLMLLEIK
		845	855	865	875	885	895
65	EMCR	DGHLWNLFVE	K-----	-FNEVTDWLK	TLKLTLSNG	LLGNCAKRFR	RVLVKLLDVY
	229E	GGKIWNQVIK	N-----	-VNSVRDWLK	SLKLNLTQQG	LLGTCAKREF	RNLGILLEAY
	PEDV	GGSIWRSFIT	G-----	-LNTMWDFCK	RLKVSFGLDG	IVVTVARKEK	RLGALLAEMY
	TGEV	YGAIWYNCMP	G-----	-CSDP-SVLG	SVQLLIGNG-	-VKVVCDCGCK	GFANQLSKGY
	OV43	GDKYYPVVVD	-DHVGLLDQA	WRVPCAG--R	RVTTFKEQPTV	KEIISMPKII	KVPYELDNDF
70	BoCoV	GDKYYPVVVD	-GHVGLLDQA	WRVPCAG--R	CVTFKEQPTV	NEIASTPKTI	KVPYELDKDF
	MHV	GQKYFPVVVD	NDTIGVLQDC	WRFPACAG--K	KVEFNKPKV	KEIPST-RKI	KINFALDATE
	AIBV	PPDVVPIKVS	-----	-IECCG--E	PWNTIFKKAY	KEPIEVDL	TVEQLLSVIY
	SARS CoV	DKEQYCALSP	--GLLATNNV	FRKLGGAPIK	GVTFG-EDTV	WEVQGY-KNV	RITFELDERV
75		905	915	925	935	945	955
	EMCR	NGFLETVCVS	VHTAGVCIKY	YAVNVV-YV	ISGFVSRVIR	RERCD--VTF	PCVSCVTTYF
	229E	NAFLDTVVST	VKIGGLTFKT	YAFDKP-YIV	IRDIVCKVEN	KTEAEWIELF	PHNDRIKSFS
	PEDV	NTYLSTVVEN	LVLGVSFSKY	YATSVV-KIV	LGGCFHSVKS	VFASV--FQI	PVQAGIEKFK
80	TGEV	NKLCNAARND	IEIGGIPFST	FKTPTNTFIE	MTDAIYSVIE	QGKALS----	-----



5	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	SLGDSGKLLS AMGDVGLCMY SLGDSVACLE HSGDAEYLLS KGDIIKVSKE KGDIIKVSKE KGDVIKVLRR EFKEFCIVNA EKDAPYMGVD	ELLKDKYTCS RLKDLHTGF SLTKDLHTLK LMLNDYSTAK VKAIEVVNPA VKAIEVVNPA VGAIEVVNPA ANEHMTGSG VITSGDITCV	ITFEMSCDCG MVMYKCSCT ITCSVVCSCG IVLAACKCGG NGHMAHGGGV NGHMAHGGGV NGRMAHAGV VAKAIADFCG VIPSCKAGGT	KKFDEQVGC SGRLEESGAV TGERIYEGCA EKEIVLERAV AKAIAVAAGQ AKAIAVAAGQ AGAIKAAGK LDFVEYCEDY TEMLSRALKK	FWIMPYTKLF LFCPTPKKAF FRMTPTLEPF FKLTPKESF QFVKETDMV QFVKETDMV SFIKETADMV VKKHGPQORL VPVDEYITTY	QKGECCICHK PYGTCNLNCA PYGACAQCAQ NYGVCSDCMQ KSKGVCATGD KSKGVCATGD KNQGVCCQVE VTPSFVKGIQ PGQGCAGYTL
10		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
15	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	1385 MQTYKLVSME PRMCTIRQLQ VLMHTFKSIV VNTCRFLSVE CYVSTGGKLC CYVSTGGKLC CYESTGGNLC CVNNVVGPRH EEAKTALKKC	1395 GTGVFVQD-- GTIIFVQKQ-- GTGIFCRD-- GSGVFVHDIL KTVLNVVGPD KTVLNVVGPD KTVLNVVGPD GDNVLHEKLV KSAFYVLPSE	1405 PAPIDIDAFP PEPVNPVSFV TTALSLDSL SKQTPEAMFV ARTQKQSYV ARTQKQSYA ARGHGKQCY AAYKNVLVDG APNAKEEILG	1415 VRPICSSVYL VKPVCSSIFR VKPLCAAFAI VKPVMHAYVT LLERVYKHLN LLERVYKHLN FLERAYOHIN VVNYVVPVLS TVSWNLREML	1425 GVKGSCHYQT GAVSCGHYQT GK-DSGHYVT GTTQNGHYMV NYDCVVTTLI NYDCVVTTLI KCDVVTTLI LGIFGVDFKM AHAEETRKLM	1435 NLYSFDKAID NIYSQNLCDV NFYDAAMAID DDIEHGYCVD SAGIFSVPST SAGIFSVPST SAGIFSVPST SIDAMREAFE PTCMDVRAIM
20		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
25	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	1445 GFGVFDIK-- GFGVNKIQF-- GYGRHQIK-- GMIKPLKKR VSLTYLLGTA VSLTYLLGTA VSLTYLLGTA GCTIRVLLFS ATIQKRYKGI	1455 ----- ----- ----- CYTSTLFINA KKQVVLVSNN KKQVVLVSNN TKNVILVSNN ----- KIQEGIVDYG	1465 -----NSSV -----WTNDAL -----YDTL NVMTRAEPK QEDFDLISK QEDFDLISK KDDFDVIEK -----LSQE VRFFFYTSKE	1475 NTVCFVDVDF NTICIKDADY NTICVKDVNW QEFKVEKVEQ QITAVEG-TK QITAVEG-TK QVTSIAG-TK HIDYFDVTCK PVASIITKLN	1485 HS-VEIEAGE NAKVEISVTP TAPLVPVADS QPIVEENKSS KLAARLSFN KLAARLSFN ALSLQAKNL OKTIYLTEDG SLNEPLVTMP	1495 VK----- IKNTVDTTPK VVEP----- IEKEEIQSPK GRSIVYETDA GRSIVYETDA CRDVKFETNA VKYR----- IGYVTHGFNL
30		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
35	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	1505 EEFVVKEKLN -----VVK ND-----DLIL NKILILN--- NKILILN--- CDSLFS----- EEAARCMR--	1515 PFAVYKNVRF AFLVHDNVAF PFYSYKNVDF PFYKAGKLSF DVAFVSTFNV DVAFVSTFNV DSCFVSSYDV SLKAPAVVSV	1525 YLGDISHLVN YQGDVDTVVN YQGDVDTVVN YQALDVLIN LQDVLSLRHD LQDVLSLRHD LQDVLSLRHD SIVLKPG SSPDVATTYN	1535 CVSFDVFNVA GVDFDFIVNA -LPCDFVNA FLEPDVIVNA IALDDDARTF IALDDDARTF IQLDVLDARF DSLQGFQGVY GYLTSSSKTS	1545 ANENLMHGGG ANENLAHGGG ANEKLSHGGG ANGDLKHMGG VQSNVDVPE VQSNVDVPE VQAHMDNLPA AKNKIVFTAD EEHFVETVSL	1555 VARAIDILTE LAKALDVYTK IAKAIQVYTK VARAIDVFTG GWRVNVKIFYQ GWRVNVKIFYQ DWRLVNFKFS DVEDKEILYV AGSYRDSWSYS
40		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
45	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	1565 GQLQSLSKDY GKLQRLSKEH GMLQKCSNDY GKLTERSNDY INGVRT-VKY INGVRP-VKY VDGVRT-VKY PTTDKSILEY GQRTGELGVEF	1575 ISSNGPLKVG IGLAGKVKVG IKAHGPIKVG LKNKSIAPG FECTGGIDIC FECFPGGIDIC FECFPGGIDIC YGLDAQRYVI LKRGDKIVYH	1585 AGVMLE--CE TGVMVE--CD RGVMLE--AL NAVFFENVIE SQDKVFGYVQ SQDKVFGYVQ SQDKVFGYVQ YLOTLAQKWN TLESPEVEHL	1595 KEN--VFNVV SLR--IFNVV GLK--VFNVV HLS--VLNAV QGFENKATVA QGSFNKATVA NGSFKVASVS VQYRDNELL DG--EVLSD	1605 GPRTG----- GPRKG----- GPRKG----- GPRNGD----- QIKALF----- QIKALF----- QIRALL----- EWRDGN--CW KLKSLLSLRE	1615 KHEHSLIVEA KHERDLLIKA KHAPELLVKA SRVEAKLCNV LDKVDILLTV LDKVDILLTV ANKVDLCTV ISSAIVLLQA VKTIKVFTTV
50		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
55	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	1625 YNSILFENGI YNTINNEQGT YKSVFANSVG YKAIKACEGK DGVNFTNRV DGVNFTNRV DGVNFTNRV AKIRFKGFLT DNTNLHTQLV	1635 PLMPLLSGCI PLTPILSCGI ALTPILSVGI ILTPILSVGI PVGESFSGSL PVGESFSGSL PVGESFSGSL EAWAKLLGGD DMSMTYGQQF	1645 FGVRIENSLK FGIKLETSL FVPLEESLS FNVRLTSLQ GNVFCDGVNV GNVFCDGVNV GNVFCDGVNV PTDFVAVCYA GPTYLDGADV	1655 ALFSCDINKP VLLDVCNTKE AFLACVGRH CLLKTVDNRG TKHKCDINYK TKHKCDINYK TKVRCSAIHK SCTAKVGDFS TKIKPHVNH	1665 LQVFVYSNE VKVFYVTDTE CKCFYGDKE LNVFVYTDQE GKVFQFQDNL GKVFQFQDNL GKVFQYSGL DANWLLANLA GKTFFVLPSD	1675 EQAVLKFLDG VCKVKDFVSG REALIKYMDG RQTIENFFS- SSEDLKAVRS SSEDLKAVRS SAADLVAVTD EHFDADYTNA DTLRSEAFEY
60		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
65	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	1685 LDLTPVID-- LVNVQKVE-- LVDAIFKEAL SFNFDQKELL SFNFDQKELL AFGFDEPQLL FLKKRVSCN-	1695 -----DVDV -----QPKI VDTTPVQEDV AYYNMLVN-- AYYNMLVN-- KYNNMLG-- -----	1705 V----- EPKPVSVIKV QQVSQKPVLP CFKWQVVVNG CSKWQVVVNG MCKWPFVVCG -----	1715 --KPFVVEGN APKPYRVDGK NFEPFRIEGA KYFTFKQANN KYFTFKQANN NYFAFKQSN -----	1725 FSFFDCG--- FSYFTED--- HAFYECNPEG NCFVNVSCLM NCFVNVSCLM NCYINVACLM -----	1735 VNALDGD-IY LLCVADDKPI LMSLGAD-KL LQSLHLTFKI LQSLHLTFKI LQSLHLTFKI -----
70		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
75	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV	1685 LDLTPVID-- LVNVQKVE-- LVDAIFKEAL SFNFDQKELL SFNFDQKELL AFGFDEPQLL FLKKRVSCN-	1695 -----DVDV -----QPKI VDTTPVQEDV AYYNMLVN-- AYYNMLVN-- KYNNMLG-- -----	1705 V----- EPKPVSVIKV QQVSQKPVLP CFKWQVVVNG CSKWQVVVNG MCKWPFVVCG -----	1715 --KPFVVEGN APKPYRVDGK NFEPFRIEGA KYFTFKQANN KYFTFKQANN NYFAFKQSN -----	1725 FSFFDCG--- FSYFTED--- HAFYECNPEG NCFVNVSCLM NCFVNVSCLM NCYINVACLM -----	1735 VNALDGD-IY LLCVADDKPI LMSLGAD-KL LQSLHLTFKI LQSLHLTFKI LQSLHLTFKI -----
80		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....



63/87

	SARS CoV	YHTLDESEFLG	RYMSALNH--	TKKWKFPQVG	GLTSIKWADN	NCYLSSVLLA	LQQLEVKFNA
		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
5	EMCR	1745	1755	1765	1775	1785	1795
	229E	LLFTNSILML	DKQGQL----	LDTKLNGILQ	QAVLDYLATV	KTVFAGNLVK	LVVE-SCTIY
	PEDV	VLFTDSMLTL	DDRGLA----	LDNALSGVLS	AAIKDCVDIN	KAIPSGNLIK	FDIG-SVVVY
	TGEV	VLFTNSNLDF	CSVGKC----	LNDVTSGLL	EAINVFKSN	KTVFAGNCVT	LDCANMISIT
10	OV43	VQWQEAWLEF	RSGRPAREFVA	LVLAKGGFKF	GDPADSRDFL	RVVFSQVDLT	GAICDFEIIAC
	BoCoV	VQWQEAWLEF	RSGRPAREFVS	LVLAKGGFKF	GDPADSRDFL	RVVFSQVDLT	GAICDFEIIAC
	MHV	WQWQEAWNEF	RSGKPLRFVS	LVLAKGSEKF	NEPSDSTDFM	RVVLEADLS	GATCDFEFVC
	AIBV						
	SARS CoV	PALQEAYYRA	RAGDAANFCA	LILAYSNTKV	GELGDVRETM	THLLQHANLE	SAKRVLNVVC
15		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	EMCR	1805	1815	1825	1835	1845	1855
	229E	M-CVVPISIND	LSFDKNLGRG	VRKLNRLKTC	VIANVPAIDV	LKKLLSSLTL	TVKFVVESNV
	PEDV	M-CVVPSEKD	KHLDDNNVQRC	TRKLNRLMCD	IVCTIPADYI	LPLVLSSLTC	NVSFVGELKA
20	TGEV	M-VVLPFDGD	ANYDKNYARA	VVKVSKLKGK	LVLAVDDATL	YSKLS--HLS	VLGFVSTPDD
	OV43	--CSIP----					
	BoCoV	K-CGVKQEQR	TGLDAVMHFG	TLSREDLEIG	YTVDCSCG--	--KKLIHCVR	DVPFLICSNT
	MHV	K-CGVKQEQR	TGVDAMHFG	TLSREDLEIG	YTVDCSCG--	--KKLIHCVR	DVPFLICSNT
	AIBV	K-CGVKQEQR	KGVDAMHFG	TLDKGDPLAKG	YTIACTCG--	--NKLHVCTQL	NVPFLICSNT
25	SARS CoV	--CGIKSYEL	RGLEACIQP-				
		KHCGQKTTTL	TGVEAVMYMG	TLSYDNLKTG	VSIPCVCGR-	--DATQYLQQ	ESSFVMMSSAP
		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	EMCR	1865	1875	1885	1895	1905	1915
30	229E	MDVNDCFKND	NVVLKITEDG	INVKDVVVS	SKSLGKQLG-	VVSDGVDSE	GVLE--INTD
	PEDV	AEA-----K	VITIKVTEDE	VNVHDVTVT	DKSFEQQVG-	VIADKDKDLS	GAVPSDLNNTS
	TGEV	VER--FYANK	SVVIKVTEDT	RSVKAVKVES	TATYGGQIG-	PCLVNDTVVT	DNKP--VVAD
	OV43	--VN-----	-----VTEDN	VNHERVSVSF	DKTYGEQLKG	TVVIKDKDVT	NQLPSAFDVG
35	BoCoV	PASVKLPKG-	VGSANIFIG-	DKVGHVHVHVK	CEQSYQLYDA	SNVKKVTDVT	GKLSDCLYLK
	MHV	PASVKLPKG-	VGSANIFIG-	DKVGHVHVHVK	CEQSYQLYDA	SNVKKVTDVT	GKLSDCLYLK
	AIBV	PEGKKLPDD-	VVAANIFTG-	GSLGHYTHVK	CKPKYQLYDA	CNVSKVSEAK	GNETDCLYLK
	SARS CoV	-----LLHFK	TQYSNCPTCG	ANNTDEVIEA	SLPYLLLFAT	DGPATVDCDE	DAVG-----
		PAEYKLQGGT	FLCANEYTG	YQCGHYTHIT	AKETLYRIDG	AHLTKMSEYK	GPVTDVDFY-K
40		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	EMCR	1925	1935	1945	1955	1965	1975
	229E	TVLSVAPEVD	WVAFYGFERA	ALFASLDVKP	YGYPNDFVGG	FRVLGTDDNN	CWVNATCIIL
	PEDV	ELLTAKAIDVD	WVEFYGFKDA	VTFATVDHSA	FAYESAVVNG	IRVLKTSDDN	CWVNATCIIL
45	TGEV	VVAKVVPNAN	WDSHYGFDKA	GEFHMLDHTG	FTFPSEVUNG	RRVIKTTDNN	CWVNATCIIL
	OV43	QKVIKAIDID	WOAHYGFERDA	AASFASSHDA	YKFEVVTSHN	FIVHKQTDNN	CWINAICLAL
	BoCoV	NLKQTFKSVL	TTYLDDVKK	IEYKPDLSQY	YCDGGKYITQ	RIIKAQFKTF	EKVDGVYTNF
	MHV	NLKQTFKSVL	TTYLDDVKK	IEYKPDLSQY	YCDGGKYITQ	RIIKAQFKTF	EKVDGVYTNF
	AIBV	NLKQTFSSKL	TTYLDDVKK	VEYNPDLSQY	YCESGKYITK	PIIKAQFRTF	EKVEGVYTNF
50	SARS CoV	TVVEVGSTNS	GHCYTQAAGQ	AFDNLAKDRK	FGKKSPLYTA	MYTFAFKNE	TSLPVAQSK
		ETSYTTTIK	VSYKLDGVTY	TEIEPKLDGY	YKKNAYYTE	QPIDLVP-TQ	PLPNASFDFN
		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	EMCR	1985	1995	2005	2015	2025	2035
55	229E	QYLKPTFKSK	GLNVLWKNFV	TGVDGPFVSF	IYFITMSSKG	QKGDAAEALS	KLSEYLISDS
	PEDV	QYSKPHFISQ	GLDAAWNKFV	LGDEIFVAF	VYVVARLMKG	DKGDAEDTLT	KLSEYLISDS
	TGEV	QFARFRFKSA	GLQAMWESYC	TGVDAMFVHW	LYWLTGVDRG	QPSDSENALN	MLSKYIVPAG
	OV43	QRLKPQWKFP	GVRGLWNEFL	ERKTQGFVHM	LYHISGVKRG	EPGDALMLH	KLGDLMNDNC
	BoCoV	KLIG--HTVC	DSLNAKLGF	SSKEFVEYKI	TEWPTATGDV	VLATDDLYVK	RYERGCTIFG
60	MHV	KLIG--HTVC	DILNAKLGF	SSKEFVEYKI	TEWPTATGDV	VLATDDLYVK	RYERGCTIFG
	AIBV	KLVG--HSIA	EKENAKLGF	CNSPFTEYKI	TEWPTATGDV	VLASDDLYVS	RYSGGCVTFG
	SARS CoV	GKSKS-VKED	VSNLATSSKA	SFDNLDTDFEQ	WYDSNIYESL	KVQESPDNFD	KYVSFTTKED
		KLTCNTRKFA	DDLNQMTGFT	KP-ASRELSV	TFFPDNLGDV	VAIDYRHYS	SFKKGAKLLH
65		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	EMCR	2045	2055	2065	2075	2085	2095
	229E	IVTLEQYSTC	DIC-----	-----KSTVV	EVKSAVVCAS	VLKDGCDVG-	-----
	PEDV	QVQLEHYSSC	VECDARF----	-----KNSVA	SINSAIVCAS	VKRDGVQVG-	-----
	TGEV	SVTIERTVTHD	GCC-----	-----CSKR	VVTAPVNVAS	VLKLGVEDG-	-----
	OV43	EIIIVTHTTAC	DKC-----	-----AKVE	KFVGPPVVAAP	LAIHGTDE--	-----
70	BoCoV	KPVIWLSHEK	ASLNSLT----	-----YFNRP	SLVDDNKFDV	LKVDVDD--	-----
	MHV	KPVIWLSHEQ	ASLNSLT----	-----YFNRP	LLVDENKFDV	LKVDVDD--	-----
	AIBV	KPVIWLGHEE	ASLNSLT----	-----YFNRP	SVVCENKFN	LPVDVSEPTD	KGPVFAAVLV
	SARS CoV	SKLPLTLKVR	GIKS-----	-----VV	DFRSKDGFIY	KLTPDFTD--	-----
		KPIVWHINQA	TTKTTFKENT	WCLRCLWSTK	PVDTSNSFEV	LAVEDTQGM	N-----
75		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	EMCR	2105	2115	2125	2135	2145	2155
	229E	---FCPHRH	KLRSRVK----	-----	-----	---FVNGRVVIT	NVGEPIISQP
	PEDV	---YCVHGI	KYYSRVK----	-----	-----	---SVRGRAIIV	SVEQLEPCAQ
80	TGEV	---LCPHGL	NYIGKVV----	-----	-----	---VVKGTIIVV	NVGKPVVAPS
		---TCVHGV	SVNVKVT----	-----	-----	---QIKGTVAIT	SLIGPIIG--

64187

OV43	----	DGGDSS	ESGAKE----	-----	-----	TKEINIIKLS	GVKKPFKVED
BoCoV	----	DGGDIS	ESDAKE----	-----	-----	PKEINIIKLS	GVKKPFKVED
MHV		TGALSGAATA	PGTAKEQKVC	ASDSVVDQVV	SGFLSDLSCA	TVDVKEVKLN	GVKKPIKVED
AIBV		-----	-----	-----	-----	-----ENSKAPVY	YPVLDAISLK
5 SARS CoV		----	LACESQ	QPTSEEVVEN	-----	-----PTIQKEVIE	CDVKTTEVVG
		..... .....	..... .....	..... .....	..... .....	..... .....	..... .....
		2165	2175	2185	2195	2205	2215
10 EMCR		SKLLNGIAYT	TFS--GSFDN	GHYVVYDAAN	NAVYDGALE	ASDLSTLAVT	AIVVVGCCVT
229E		SRLLSGVAYT	AFS--GPVDK	GHYTVYDTAK	KSMYDGDRE	KHDLSSLSTV	SVVMVGGYVA
PEDV		HLFLKGVSYT	TFLDNGNGV	GHYTVFDHGT	GMVHDGDAFV	PGDLNVSPVT	NVVVSEQTAV
TGEV		-EVLEATGYI	CYS--GSNRN	GHYTYDNRN	GLVVDAAKAY	HFNRDLLOQT	TAIASNFVVK
OV43		SVIVNDDTSE	TKYVKSLSIV	DVYDMWLTGC	KYVVRTANAL	SRVNVPTIR	KFIKEGMTLV
BoCoV		SVIVNDDTSE	IKYVKSLSIV	DVYDMWLTGC	RCVVRTANAL	SRVNVPTIR	KFIKEGMTLV
15 MHV		SVVVNDPTSE	TKVVKLSIV	DVYDMFLTGC	RYVVMANEL	SRLVNSPTVR	EYVKWGMTKI
AIBV		AIWVEGNANF	VVG-----HP	NYYSKSLHIP	TFWENAENFV	KMGDKIGGVT	MGLWRAEHLN
5 SARS CoV		NVILKPSDEG	VKVTPQLGHE	DLMAAYVENT	SITIKKPNE	SLALGLKTIA	THGIAAINS
		..... .....	..... .....	..... .....	..... .....	..... .....	..... .....
		2225	2235	2245	2255	2265	2275
20 EMCR		S-----	-----NVPP	IVSEKISVMD	KLDTG---AQ	KFFQFGDFVM	NNIVLELTWL
229E		-----	-----PV	NTVKPKPVIN	QLDEK---AQ	KFFDFGDFLI	HNFIFFFTWL
PEDV		V-----	-----IKDP	VKKAELDATK	LLDTMNYASE	RFFSFGDFMS	RNLITVFLYI
TGEV		KPOAERPKN	CAFNKVAASP	KIVQEQKLLA	IESGANYALT	EFGRYADMFF	MAGDKILRL
25 OV43		SIP-----	-----IDLL	NLREIKPAVN	VVKAVRNKIS	VCFNFIKWLF	VLLFGWIKIS
BoCoV		SIP-----	-----IDLL	NLREIKPVFN	VVKAVRNKIS	ACFNFIKWLF	VLLFGWIKIS
MHV		VIP-----	-----AKLV	LLRDEKQEFV	APKVVAKAVI	ACYSAVKWFF	LYCFSWIKFN
AIBV		KPN-----	-----LERI	FNIAKKAIVG	SSVVTTCGK	LIGKAATFIA	DKVGGGVVRN
30 SARS CoV		PWS-----	-----KILA	YVKPFLG---	---QAATTSN	CAKRLAQRVF	NNYMPYVFTL
		..... .....	..... .....	..... .....	..... .....	..... .....	..... .....
		2285	2295	2305	2315	2325	2335
35 EMCR		LSMFSLRLTS	IMKHDIKIVA	KAPKRTGVIL	TRSFKYNIRS	ALFVVQKQWC	VIVTLFKELL
229E		LSMFTLCKTA	VITGDVKIMA	KAPQRTGVVL	KRSLKYNLKA	SAAVLKSXWW	LLAKFTKLL
PEDV		LSILGLCFRA	FRKRDVKVLA	GVPQRTGIIL	RKSMRYNAKA	LGVEFFKLKLY	WPKVLGKFS
TGEV		LEVFKYLLVL	FMCLRSTKMP	KYVKVP-PLA	FKDFGAKVRT	LNMYRQLNKP	SVWRYAKLV
OV43		ADNKVIYTTT	IASKLTCKLV	ALAFKNAFLT	FKWSMVARGA	CIIATIFLLW	FNFIYANVIF
BoCoV		ADNKVIYTTT	VASKLTCKLV	ALAFKNAFLT	FKWSMVARGA	CIIATIFLLW	FNFIYANVIF
40 MHV		TDNKVIYTTT	VASKLTCKLV	CLAFKNAFLT	FNWNVVSRGF	FLVATVFLW	FNFIYANVIL
AIBV		ITDSIRGLCG	ITRGHFERMK	SQFLKTLMF	FLFYFLKASV	KSVVASYKTV	LCKVVLATLL
5 SARS CoV		LFQLCTFTKS	TNSRIRASLP	TTIAKNSVKS	VAKLCLDAGI	NYVKSPPFSK	LFTIAMWLLL
		..... .....	..... .....	..... .....	..... .....	..... .....	..... .....
		2345	2355	2365	2375	2385	2395
45 EMCR		LLYAIYALVF	MIVQFSPFNS	LLCGDIVSGY	EKSTFN----	--KDIYCGNS	MVCKMCLFSY
229E		LIYTLYSVVL	LCVRFPGPFN	-FCSETVNGY	AKSNFV----	--KDDYCDGS	LGCKMCLFGY
PEDV		GIYALYALLF	MTIREFTPIGS	PVCDDVVAGY	ANSSFD----	--KNEYCN-S	VICKVCLYGY
TGEV		LLIAIYNFFY	LFVSIPVVKH	LTGNGAVQAY	KNSSFI----	--KSAVCGNS	ILCKACLASY
50 OV43		SDFYLPKIGF	LPTFVGKIAQ	WIKNTFSLVT	ICDLYSMQDV	GFKNQYCNCS	IACQFCLAGF
BoCoV		SDFYLPKIGF	LPTFVGKIAQ	WIKNTFSLVT	ICDLYSIQDV	GFKNQYCNCS	IACQFCLAGF
MHV		SDFYLPNIGF	FPTFVGQIVA	WVKTFFGIFT	LCDLYQVSDV	GYRSSFCNGS	MVCELCSFGF
AIBV		IVFEVYTSNP	VMFTGIRVLD	FLFEGSLCGP	YKDYGK--DS	FDVLYRCADD	FTCRVCLHDK
5 SARS CoV		LSICLGSLLC	VTAAGVLLS	NFGAPSYCNG	VRELYLNSSN	VTTMDFCEGS	FPCSICLSGL
		..... .....	..... .....	..... .....	..... .....	..... .....	..... .....
		2405	2415	2425	2435	2445	2455
55 EMCR		QEFNDLDHTS	LVWKHIR---	-----D--P--	-ILISLQPFV	ILVILLIFGN	MYLRFGLLYF
229E		QELSQFSHLD	VVWKHIT---	-----D--P--	-LFSNMOPFI	VMVLLIFGD	NYLRCFLLYF
PEDV		QELSDFSHTQ	VVWQHRL---	-----D--P--	-LIGNVMPFF	YLAFLAIFGG	VYVKAITLYF
TGEV		DELADFHQLQ	VTWDFKS---	-----D--P--	-LWNRLVQLS	YFAFLAVFGN	NYVRCLMYF
60 OV43		DMLDNYKAID	VVQYEADRRR	FVDYTGVLKI	VIELIVSYAL	YTAWFYPLFA	LISIQLTTW
BoCoV		DMLDNYKAID	VVQYEADRRR	FVDYTGVLKI	VIELIVSYAL	YTAWFYPLFA	LISIQLTTW
MHV		DMLDNYDAIN	VVQHVVDRRV	SFDYISLFLK	VVELVIGYSL	YTVCYFPLFG	LIGMQLTTW
AIBV		DSLHLKHYAY	SVEQVYKDA	SG-----	--FIFNWNWL	YLVFLILFVK	PVAGFVICY
65 SARS CoV		DSLDSYPAL	TIQVTIS--S	YKLDLTILGL	AAEWVLAJML	FTKFFYLLGL	SAIMQVFFGY
		..... .....	..... .....	..... .....	..... .....	..... .....	..... .....
		2465	2475	2485	2495	2505	2515
70 EMCR		VAQFISTFG-	-SFLGFHQKQ	WELHVFVPDV	LCNEFLATFI	VCKIVLEVRH	IIVGCNNADC
229E		VAQMISTVG-	-VFLGYKETN	WELHFIPEDV	ICDELLVTVI	VIKVISEVRH	VLFGCENPDC
PEDV		IFQYLNLSG-	-VFLGLQSQI	WELQLVFPDV	FGDEIVVFFI	VTRVLMFIKH	VCLGCKRASC
TGEV		VSQYLNLSG-	-SYFGYVEYS	WELHVVNFES	ISAEFVIVVI	VVKAVLALKH	IVFACSNPSC
75 OV43		LPELFMLST-	-LHWSVRLLV	SLANMLPAHV	FMRFYIIIAS	FIKLFSLFRH	VAYGCSKSGC
BoCoV		LPELFMLST-	-LHWSVRLLV	SLANMLPAHV	FMRFYIIIAS	FIKLFSLFRH	VAYGCSKSGC
MHV		LPEFFMLET-	-MHWSARFFV	FVANMLPAFT	LLRFYIVVTA	MYKIFCLCRH	VMYGCSRPSC
AIBV		CVKYLVLNST	VLQTVGVCFLD	WFTQTVFSHF	NFMGAGFYFW	LFYKIYQVH	HILYCKDVTC
5 SARS CoV		FASHFISN--	--SWLMWFII	SIVQMAPVSA	MVRMYIFTAS	FYIWKSYVH	IMDGCTSSTC
		..... .....	..... .....	..... .....	..... .....	..... .....	..... .....
		2525	2535	2545	2555	2565	2575
80							



65/87

5	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	VACSKSARLK IACSKSARLK VACSKSARLK KTCSTRARQT LFCYKRNRSL LFCYKRNRSL LFCYKRNRSL EVCKRVARSN MMCYKRNRAT	RVPLQTIING RFPVNTIIVNG RVPVQTIIFQG RPIQVIVVNG RVKCSSTIVGG RVKCSSTIVGG RVKCSSTIVGG RQEVSVVVG RVECTTIVNG	MHKSFYVNNAN VQSRFYVNNAN TSKSFYVHAN SMKTYYVHAN MIRYYDVMAN MIRYYDVMAN MIRYYDVMAN RKQIVHVYTN MKRSFYVYAN	GGTCFCNKHN GGSKFCKKHN GGSKFCKKHN GTGKFCKKHN GGTGFCSKHQ GGTGFCSKHQ GGTGFCAKHQ SGYNFCKRHN GGRGFCCKTHN	FFCVNCDSEFG FFCVDCDSYG FFCLNCDSEYG FYCKNCDSEYG WNCIDCDSEYG WNCIDCDSEYG WNCIDCDSEYG WYCRNCDSEYG WNCINCDSEYG	PGNTFINDGI YGSTFITPEV PGCTFINDVI FENTFICDEI PGNTFITVEA PGNTFITVEA PGNTFITVEA HONTFMSPEV TGSTFISDEV
10		2585	2595	2605	2615	2625	2635
15	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	ARELGNVVKT SRELGNITKT ATEVGNVVKL VRDLNSNVKQ ALDLSKELKR ALDLSKELKR AADLSKELKR AGELSEKLR ARDLSLQFKR	AVQPTAPAYV NVQPTGPAYV NVQPTGPATV TVYATDRSHQ PIQPTDVAYH PIQPTDVAYH PVNPTDSAYY HVKPTAYAYH PINPTDQSSY	IDKELTAVNG MIDKVEFENG LIDKVEFSNG EVTKEVCSHG TVTDVKQVGC TVTDVKQVGC LVTEVKQVGC VDEACLVDD IVDSVAVKNG	FYRLYSGLTF FYRLYSCTTF FYRLYSGLTF FYRFYVGDFF SMRLFYDRDG SMRLFYDRDG SMRLFYERDG FVNLKYKAAT ALHLYFDKAG	WRYDFEDITES WRYNEDITES WRYNEDITES TSYDYDVVHKH QRTYDDVNAS QRTYDDVNAS QRTYDDVNAS PGKDSASSAV QKTYERHPLS	KYSCKE---- KYSCKE---- KYSCKE---- KYSCKE---- LFVDYSNLLH LFVDYSNLLH LFVDYSNLLH KCFSVTDFLK HFVNLNLDLRA
20		2645	2655	2665	2675	2685	2695
25	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	-VLKNCNVLE -VFKNCNVLD -ALKNCIIIT -VLKSMILLD SKVKSVPNMH SKVKSVPNMH SKVKSVPNMH KAVFLKEALK NNTKGSLLPIN	NEFIVYNN-- DFIVFNN-- DFIVFNN-- DFIVYSP-- VVVVEN-- VVVVEN-- VVVVEN-- CEQISNDGFI VIVFDGK--	-----SGSNI -----NGTNV -----NGSNV -----SGSAL -----DADK -----DADK -----DADK VCNTQSAHAL -----SKCDE	TQIKNACVYF TQVKNASVYF NQVKNACVYF ANVRNACVYF ANFLNAAVFY ANFLNAAVFY AGFLNAAVFY BEAKNAAIYY SASKSASVYY	SQQLCEPIKL SQQLCRPIKL SQMLCKPVKL SQLGKPKIKI AQSLFRPILM AQSLFRPILM AQSLFRPILM AQYLCKPILI SQLMCQPIKL	VNSELSTLS VDSSELLSTLS VDSALLASLS VNSDLLELDS VDKNLITTTAN VDKNLITTTAN VEKKLITTTAN LDQALYEQLV LDQVLVSDVG
30		2705	2715	2725	2735	2745	2755
35	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	--VDFNGVLH --VDFNGVLH --VDFGASLH --VDFKGALF TGTSVTETMF TGTSVTETMF TGLSVSQTMF V-EPVSKSVI DSTEVSVMKF	KAYVDVLCNS KAYIDVLRNS SAFVSVLNS NAKKNVKNKS DVYVDTFLSM DVYVDTFLSM DLYVDSLLGV DKVCSILSSI DAYVDTFSAT	FFKELTANMS FGKDLNANMS FGKDLSSCND FNVDSSECKN FDVDKKSLSNA FDVDKKSLSNA LDVDRKSLTS ISVDTAALNY FSVPMEKLKA	MAECKATLGL LAECKRALGL MQDCKSTLGF LDECYRACNL LIATAHSSIK LIATAHSSIK FVNAAHNSLK KAGTLRDALL LVATAHSELA	T----- S----- DD----- N----- QGTQIYKVL QGTQIYKVL EGVQLEQVMD S----- KGVALDGVLS	----- ----- ----- ----- TFLSCARKSC TFLSCARKSC TFIGCARRKC ----- TFVSAARQG-
40		2765	2775	2785	2795	2805	2815
45	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	-----VSDDD -----ISDHE -----VPLDT -----VSFST SIDSVDVTKC SIDSVDVTKC AIDSVDVTKS -----ITKDEE VVDTDVDTKD	FVSAVANAHR FTSAISNAHR FNAVAEAAHR FEMAVNNAHR LADSVMSAVS LADSVMSAVS ITKSIMSAYN AVDMAIFCHN VIECLKLSHH	YDVLSDLSF CDVLSDLSF YDVLTDMSF FGILITDRSF AGLELTDESC AGLELTDESC AGVDFDESC HDVDTGDDG SDLEVTGDDG	NNFFISYAKP NNFVSSYAKP NNFTTSYAKP NNFWPSKVVP NNLVPTYLKS NNLVPTYLKG NNLVPTYVKS TNVIPSYGID NNFMLTYNKV	EDK-LSVYDI EEK-LSAYDL EEK-FPVHDI GSSGVSAMDI DN--IVAADL DN--IVAADL DT--IVAADL TG-KLTPRDR EN--MTPRDL	ACCMRAGSKV ACCMRAGAKV ATCMRVGAKI GCMTSDAKI GVLIONSNAH GVLIONSNAH GVLIONSNAH GFLINADASI GACIDCNARH
50		2825	2835	2845	2855	2865	2875
55	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	VNHNVLKES VNANVLTKDQ VNHNVLVKDS VNAKVLTKQ VQGNVAKIAG VQGNVAKIAG VQANVAKAAN ANLRVKN--A INAQVAKSHN	IPIVWGVKDF TPIVWHAKEF IPVWVLVRDF KSVVWLSQDF VSCIWSVDAF VSCIWSVDAF VACIWSVDAF PPVWVKFSEL VSLIWNVKDY	NTLSQEGKRY NSLSAEGKRY IALSEETRY AALSSTAQKV NQFSDFQHK NQFSDFQHK NQFSDFQHK IKLSDSLKY MSLSEQLRQ	LVKTTKAKGL IVKTSKAKGL IIRTTKVKGI LVKTFVEEGV LKKACCKTGL LKKACCKTGL LKKACCKTGL LISATVKSGV IRSAAKNNNI	TFLLTFNDNQ TFLLTFNDNQ TFLLTFNDNQ TFLLTFNDNQ KLKLTYNKQM KLKLTYNKQM KLKLTYNKQM RFFITKSGAK PFRITCATTR	AITQVP---- AVTQIP---- MHTTIP---- SDDDLPEYERF ANVSFLT--- ANVSFLT--- ANVSFLT--- QVIACHT--- QVNVNIT---
60		2885	2895	2905	2915	2925	2935
65	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	ATSIVAKQGA ATSIVAKQGA TVCIANKKGA TESVSPKSGS --TPFSLKGG --TPFSLKGG --TPFSLKGG --TPFSLKGG --TPFSLKGG	G----- GD----- GLPS----- G----- A--V----- A--V----- A--V----- A--V----- A--V-----	--FKRTYNEL --AGHSLTWL --FSKVKKFF --FDDVITQL --FSYFVYVC --FSYFVYVC --FSYFVYVC --FSYFVYVC --FSYFVYVC	WYVCLFVVAL WLLCGLVCLI WFLCLFVAA KQIVILVVFV FVLSLVCFIG FVLSLVCFIG FVLSLVCFIG FVNLICFIV YILFTACCSG	FIGVSFID-- QFYLCEFFMPY FFALSFLD-- IFICGLCSVY --LWCLMPT --LWCLMPT --LWCLMPT --LWCLMPT --LWCLMPT	---TITTVTS ---FMYDIVSS ---ESTQVSS SVATQSYIES YTVHKSDQFL YTVHKSDQFL YTVHKSDQFL YTVHKSDQFL YTVHKSDQFL
70		2985	2995	3005	3015	3025	3035
75	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV	QRLLEKKGAG GIVSGTFKCF KSYFKWLLIF YILFTACCSG YYMEVSKSF VHPMYDVNST					

66/07

SARS CoV		--TKISLKGK K--I----- --VSTCFKLM LKATLLCVLA ALVCYIVMPV HTLSIHGYS					
		2945	2955	2965	2975	2985	2995
5	EMCR	FHG YDFK YIE	NGQLKVFEAP	LHCVRNVFDN	FNQWHEAKFG	VVTNSDKCP	IVVG----VS
	229E	FEGYDFK YIE	NGQLKNFEAP	LKCVNRNVFEN	FEDWHYAKFG	FTPLNKQSCP	IVVG----VS
	PEDV	DSDYDFK YIE	SGQLKTFDNP	LSCVHNVFEN	FDQWHDAKFG	FTPVNPNSCP	IVVG----VS
	TGEV	AEGYDYMVIK	NGIVQPFDDT	ISCVHNTYKG	FGDWFKAKYG	FIPTEGKSCP	IVVGT--VFDL
	OV43	PVYASYKVLD	NGVIRDVSVE	DVCFANKFEQ	FDQWYESTFG	LSYYSNSMAC	PIVVA-VIDQ
10	BoCoV	PVYASYKVLD	NGVIRDVSVE	DVCFANKFEQ	FDQWYESTFG	LSYYSNSMAC	PIVVA-VVDQ
	MHV	PLYASFVID	NGVLRDVTVT	DACFANKFIQ	FDQWYESTFG	LVYYRNSRAC	PVVVA-VIDQ
	AIBV	LHVEGFKVID	KGVLREIVPE	DTCFSNKFFVN	FDAFWGRPYD	NSRNCPIVTA	VIDDGDGTAT
	SARS CoV	NEIGYKAIQ	DGVTRDIIST	DDCFANKHAG	FDAWFSQRGG	S--YKNDKSC	PVVAA--IITR
15		3005	3015	3025	3035	3045	3055
	EMCR	ERINVPVGPV	TNVYLVG---	-KTLVFTLQA	AFGNTGVCYD	FDGVTTS---	--DKCIFNSA
	229E	EIVNTVAGIP	SNVYLVG---	-KTLIFTLQA	AFGNAGVCYD	IFGVTTT---	--EKCIPTSA
	PEDV	DEARTVPGIP	AGVYLAG---	-KTLVFAINT	IFGTSGLCFD	ASGVADK---	--GACIFNSA
	TGEV	ENMRPIPDVP	AYVSVIG---	-RSLVFAINA	AFGVTNMCYD	HTGNAVSKDS	YFDTCVFNTA
	OV43	DFGSTVENVP	TKVLRYG---	YHVLHFITHA	LSADGVQCYT	PHSQISYSNF	YASGCVLSSA
	BoCoV	DFGSTVENVP	TKVLRYG---	YHVLHFITHA	LSADGVQCYT	PHSQISYSNF	YASGCVLSSA
	MHV	DIGYTLENVP	TKVLRYG---	FHVLHFITHA	FATDSVQCYT	PHMQIPYDNF	YASGCVLSSL
	AIBV	GVPGFVSWVM	DGVMFHMTQ	TERKPYIPT	WFNREIVGYT	QDSIITEGSF	YTSIALFSAR
25	SARS CoV	EIGFIVEGLP	GTVLRAIN--	GDFLHFLPRV	FSAVGNICYT	PSKLIYSDF	ATSACVLAAE
30		3065	3075	3085	3095	3105	3115
	EMCR	CTRLEGLGGD	-NVYCYNTDL	IEGSKPYSIL	QPNAYYKYDV	K-NYVRFPEI	LARGFGLRTI
	229E	CTRLEGLGGN	-NVYCYNTAL	MEGSLPYSSI	QANAYYKYDN	G-NFIKLEPV	IAQGFGRFTV
	PEDV	CTTLTGLGGT	-AVYCYKNGL	VEGAKLYSEL	APHSYKMDV	G-NAVSLPEI	ISRGFGIRTI
	TGEV	CTTLTGLGGT	-IVYCAKQGL	VEGAKLYSDL	MPDYVYEHAS	G-NMVKLPAI	IR-GLGLRFV
	OV43	CTMFTMADGS	PQPYCYTEGL	MQNASLYSSL	VPHVRYNLAN	AKGFIRFPEV	LREGL-VRIV
	BoCoV	CTMFTMADGS	PQPYCYTEGL	MQNASLYSSL	VPHVRYNLAN	AKGFIRFPEV	LREGL-VRIV
	MHV	CTMLAHADGT	PHPYCYTEGI	MHNASLYDSL	APHVRYNLAN	SGYIRFPEV	VSEGI-VRIV
	AIBV	CLYLTASNTF	QLYCFNGDND	APGALPFGSI	IPHRVYFQPN	GVRLIVPQOI	LHTPY---VV
	SARS CoV	CTIFKIDAMGK	PVPYCYDTNL	LEGSISYSEL	RPDTRYVLM	G-SIIQFPNT	YLEG-VRV
40		3125	3135	3145	3155	3165	3175
	EMCR	RTLATRYCRV	GECRDSHKG	CFGFDKQWVN	DGRVD---DG	YICGDGLIDL	LVNVLSIFSS
	229E	RTIATKYCRV	GECVESNAGV	CFGFDKQWVN	DGRVA---NG	YVCGTGLWNL	VFNILMSFSS
	PEDV	RTKAMTYCRV	GOCVQSAEGV	CFGADRFVY	NAESG---SD	FVCGTGLFTL	LMNVISVFSK
	TGEV	KTOATTYCRV	GECIDSKAGF	CFGDNWVFV	DNEFG---NG	YICGNSVLGF	EKNVFKLFNS
	OV43	RTRMSYCRV	GLCEEADDEGI	CFNFNGSWVL	NNDYRSLPG	TFCGRDVFDL	IYQLFKGLAQ
	BoCoV	RTRMSYCRV	GLCEEADDEGI	CFNFNGSWVL	NNDYRSLPG	TFCGRDVFDL	IYQLFKGLAQ
	MHV	RTRSMYCRV	GLCEDAEAGV	CFNFNGSWVL	NNPYRAMPG	TFCGRNADF	IHQVLGGLVR
	AIBV	KFVSDSYCRG	SVCEYTRPGY	CVSLNPQWVL	ENDEYTSKPG	VFCGSTVREL	MFSMVSTFFT
	SARS CoV	TTFDAEYCRH	GTCERSEVGI	CLSTSGRWVL	NNEHYRALS	VFCGVDAMNL	IANIFTPLVQ
50		3185	3195	3205	3215	3225	3235
	EMCR	SFSVVAAMSGH	MLFNFLFAAF	ITFLCFVLTK	FKRVFGDLSY	GVFTVVCATL	INNISYVVTQ
	229E	SFSVVAAMSGQ	ILLNLCALGAF	AIFCCFLVTK	FRRMFGDLSV	GVCTVVAVL	LNNVSYIVTQ
	PEDV	TVPVTVLSGQ	ILFNCLIAFV	AVAVCLFLTK	FKRMFGDMSV	GVFTVGACTL	LNNVSYIVTQ
	TGEV	NMSVVAATSGA	MLVNIIAACL	AIAMCYGVLK	EKKIFGDCFT	LIVMIIVTLV	VNNVSYIVTQ
	OV43	PVDFLALTAS	SIAGAILAVI	VVLGFYLIK	LKRAFGDYTS	IVFVNVIVWC	VNFMMLEFVQ
	BoCoV	PVDFLALTAS	SIAGAILAVI	VVLGFYLIK	LKRAFGDYTS	IVFVNVIVWC	VNFMMLEFVQ
	MHV	PIDFFALTAS	SVAGAILAI	VVLAFYLIK	LKRAFGDYTS	VVIVNVIVWC	INFLMLFVQ
	AIBV	GVN-PNIYMQ	LATMFLILV	VVLIFAMVIK	FQGVFKAYAT	TVFITMLVWV	INAFILCVHS
60	SARS CoV	PVGALDVSAS	VVAGGITAIL	VTCAAYYFMK	FRRVFGYENH	VVAANALLFL	MSFTILCLVP
65		3245	3255	3265	3275	3285	3295
	EMCR	N-LFFMLLYA	ILYFVFTRTV	R--YAWIWHI	AYIVAYFLLI	PWWLLTWFSF	AAFLELLPNV
	229E	N-LVTMIAYA	ILYFFATRSL	R--YAWIWCA	AYLIAYISFA	PWWLCAWYEL	AMLTGLLPSL
	PEDV	N-TLGMGLYA	TLYFLCTKGV	R--YMWIWHL	GFLISYILIA	PWWVLMVYAF	SAIFEFMPNL
	TGEV	N-TFFMIIYA	IVYFYTITKL	A--YFGILDA	GPIIAYINMA	PWYVITAYIL	VFLYDSLPSL
	OV43	VYPILSCVYA	ICYFYATLYF	PSEISVIMHL	QWLVMYGTIM	PLWFCLLYIA	VVVSNAHAFV
	BoCoV	VYPILSCVYA	ICYFYATLYF	PSEISVIMHL	QWLVMYGTIM	PLWFCLLYIS	VVVSNAHAFV
	MHV	VYPTLSCLYA	CFYFYTTLYF	PSEISVIMHL	QWLVMYGTIM	PLWFCLLYIA	VVVSNAHALW
	AIBV	YNSVLAVILL	VLYCYASLVT	SRNTFYIMHC	WLVTFTGLIV	PTWLACCYLG	FIIXMYTPLF
	SARS CoV	AYSFLPGVYS	VFLYLTFTYF	TNDVSFLAHL	QWFAMFSPIV	PFWITAIYVF	CISLKHCHWF
75		3305	3315	3325	3335	3345	3355
	EMCR	FKLKISTQ--	--LFEGDKFI	GTFFESAAAGT	FVLDMRSYER	LINT--ISPE	KLKNYAASYN
	229E	LKLKVSTN--	--LFEGDKFV	GTFFESAAAGT	FVIDMRSYER	LANS--ISPE	KLKSYAASYN
	PEDV	FKLKISTQ--	--LFEGDKFV	GTFFESAAAGT	FVLDMHAYER	LANS--ISTE	KLRQYASTYN
80	TGEV	FKLKVSTN--	--LFEGDKFV	GNFESAAAGT	FVIDMRSYET	IVNS--TSIA	RIKSYANSFN

	OV43	FSYCRKLG--	--TSVR--SD	GTFEEMALTT	FMTTKDSYCK	LKNS---LSDV	AFNRYLSLYN
	BoCoV	FSYCRLQG--	--TSVR--SD	GTFEEMALTT	FMTTKDSYCK	LKNS---LSDV	AFNRYLSLYN
	MHV	FSYCRKLG--	--TEVR--SD	GTFEEMSITT	FMTTKDSYCK	LKNS---VSDV	AFNRYLSLYN
5	AIBV	LWCYGTTKNT	RKLVDGNEFV	GNYDLAASKT	FVIRGSEFYK	LTNE---IGD	KFEAYLSAYA
	SARS CoV	FNNYLRKR--	--VMFNGVT	STFEAAALCT	FLLNKEMYLK	LRSETLLPLT	QYNRYLAYLN
		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		3365	3375	3385	3395	3405	3415
10	EMCR	KYKYYSGSAS	EADYRCACYA	HAKAMLDYA	-KDHNMDLYS	PPTISYN-ST	LQSGLKKMAQ
	229E	RYKYYSGNAN	EADYRCACYA	YLAKAMDFA	-RDHNDILYT	PPTVSYG-ST	LOAGLRKMAQ
	PEDV	KYKYYSGSAS	EADYRLACFA	HAKAMMDYA	-SNHNDTLYT	PPTVSYN-ST	LOAGLRKMAQ
	TGEV	KYKYITGSMG	EADYRMACYA	HKGKALMDYS	-VNRTDMLYT	PPTVSVN-ST	LQSGLRKMAQ
	OV43	KYRYXSGKMD	TAAYREAACS	QLAKAMDTFT	NNNGSDVLQY	PPTASVSTSF	LQSGIVKMVN
15	BoCoV	KYRYXSGKMD	TAAYREAACS	QLAKAMDTFT	NNNGSDVLQY	PPTASVSTSF	LQSGIVKMVN
	MHV	KYRYFSGKMD	TAAYREAACS	QLAKAMDTFT	NNNGSDVLQY	PPTASVSTSF	LQSGIVKMVN
	AIBV	RLKYXSGTGS	EQDYQLQACRA	WLAYALDQYR	-NSGVEIYVT	PPRYSIGVSR	LQSGFKKLVS
	SARS CoV	KYKYFSGALD	TTSYREAACC	HAKALNDFS	-NSGADVLYQ	PPQTSITSAT	LQSGFRKMAF
		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		3425	3435	3445	3455	3465	3475
20	EMCR	PSGCVERCVV	RVCYGSTVLN	GVWLGDTVTC	PRHVIAPSTT	VL-IDYDHAY	STMRLHNFSV
	229E	PSGFVEKCVV	RVCYGNTVLN	GLWLGDIVYC	PRHVIASNTT	SA-IDYDHEY	SIMRLHNFSI
	PEDV	PSGVVEKCI	RVCYGNMALN	GLWLGDIVMC	PRHVIASSTT	ST-IDYDYAL	SVLRLHNFSI
25	TGEV	PSGLVEPCIV	RVSYGNLVNL	GLWLGDDEVIC	PRHVIASDTT	RV-INYENEM	SSVRLHNFSI
	OV43	PTSKVEPCVV	SVTYGNMTRLN	GLWLDDKVYC	PRHVICASD	MTNPDPYTNLL	CRVTSSDFTV
	BoCoV	PTSKVEPCIV	SVTYGNMTRLN	GLWLDDKVYC	PRHVICASD	MTNPDPYTNLL	CRVTSSDFTV
	MHV	PTSKVEPCVV	SVTYGNMTRLN	GLWLDDKVYC	PRHVICASD	MTNPDPYTNLL	CRVTSSDFTV
	AIBV	PSSAVEKCI	SVSYRGNNLN	GLWLGDIVYC	PRHVICASSD	MTNPDPYTNLL	CRVTSSDFTV
30	SARS CoV	PSGKVEGCMV	QVTCGTTLN	GLWLDDTVYC	PRHVLGKFSG	DQ---WNVDL	NLANNHDFEV
		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		3485	3495	3505	3515	3525	3535
35	EMCR	SHNG-VFLGV	VGVTMHGSQL	RKIKVSQNVH	TPKHVFRTLK	PGASFNILAC	YEGIASGVFG
	229E	ISGT-AFLGV	VGATMHGVTL	KIKVSQTNMH	TPRHVSFTLK	SGEGFNILAC	YDCACQGVFG
	PEDV	SSGN-VFLGV	VSATMRGALL	QIKVQNQNVH	TPKYTYRTVR	PGESFNILAC	YDGAAAGVYG
	TGEV	SKNN-VFLGV	VSARYKGVNL	VLKVNQNVNH	TPEHKFSKIK	AGESFNILAC	YEGCPSGVYG
	OV43	LFDR-LSLTV	MSYQMGRGCL	VLTVTLQNSR	TPKYTFGVVK	PGETFTVLAA	YNGKPQGAFF
40	BoCoV	LFDR-LSLTV	MSYQMGRGCL	VLTVTLQNSR	TPKYTFGVVK	PGETFTVLAA	YNGKPQGAFF
	MHV	MSGR-MSLTV	MSYQMGGSL	VLTVTLQNPV	TPKYTFGVVK	PGETFTVLAA	YNGKSQGAFF
	AIBV	TTQHGVTLNV	VSRLKGAVAL	ILQTAVANAE	TPKYKPIKAN	CGDSFTIACA	YGGTVVGLYP
	SARS CoV	QAGN-VQLRV	IGHSMQNCCL	RLKVDTSNPK	TPKYKEFVRIQ	PGQTFSVLIAC	YNGSPSGVYQ
		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		3545	3555	3565	3575	3585	3595
45	EMCR	VNLRTNFITK	GSFINGACGS	PGYNVRNDGT	VEFCYLHQIE	LGSGAHVGSD	FTGSGVGNFD
	229E	VNMRTNWITR	GSFINGACGS	PGYNLNK-GE	VEFVYMHOIE	LGSGSHVGSS	FDGVMYGGFE
	PEDV	VNMRSNYTIR	GSFINGACGS	PGYNINN-GT	VEFCYLHQLE	LGSGCHVGSD	LDGVMYGGYE
50	TGEV	VNMRSQGTIK	GSFIAGTCGS	VGYYLEN-GI	LYFVYMHMLE	LGNGSHVGSN	FBGMYGGYE
	OV43	VTMRSSYTIK	GSFLCGSCGS	VGYYIMG-DC	VKFFVYMHOLE	LSTGCHTGTD	FNGDFYGPYK
	BoCoV	VTMRSSYTIK	GSFLCGSCGS	VGYYIMG-DC	VKFFVYMHOLE	LSTGCHTGTD	FNGDFYGPYK
	MHV	VTMRSSYTIK	GSFLCGSCGS	VGYYVLTG-DS	VRFVYMHOLE	LSTGCHTGTD	FNGDFYGPYK
	AIBV	VTMRSSYTIK	GSFLCGSCGS	VGYYVLTG-DS	VRFVYMHOLE	LSTGCHTGTD	FNGDFYGPYK
	SARS CoV	CAMPNPHITK	ASFFLAGACGS	VGFINIEK-GV	VNFFVYMHLE	LPNALHTGTD	LMGEFYGGYV
		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		3605	3615	3625	3635	3645	3655
55	EMCR	DQPSLQVES	NLMUSDNVVA	FLYAALLNGC	R-----WWL	RSTRVNVDFG	NEWAMANGYT
	229E	DQPNLQVES	NQMLTVNVVA	FLYAAILNGC	T-----WWL	KGEKLEVAHY	

5	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	YGVNLQS--- GKVIFGLKTM FLFSVFETMF WAELFIYTNT IWPNPVILT IPCLLLFLSL FGVNLQS--- GKTTSMEKSI SLFAGFFVMF WAELFVYTTT IWVNPGLTLP EMILLVALSL YGVNLQG--- GYVSRACRNV LLVGSFLTFF WSELVSYTKF FWVNPGLVTP MFACLSLLSS YGVNLQA--- GKVKSFYPI MTAMTILFAF WLEFFMYTTF TWINPTFVSI VLAVTTLIST AGIKLQSKRT RLKFGTVCWI MASTFLFSCI ITAFVKWTFM MYVTNNMFS- ITFCALCVIS AGIKLQSKRT RLKVGIVCWI MASTFLFSCI ITAFVKWTFM MYVTNNMFS- ITFCALCVIS AGVKLQSKRT RVVKGTCWII LASTLLFCSI ISAFVKWTFM MYVTNNMFS- ITFCALCVIS GGVRLQS--- -SFVRKATSW FWSRCVLACF LFLVLCVLF TAVPLKFVY AAVILLMAVL SGVTFQKFK KIVKGTHHWM LLTFLTSLLI LVQSTQWSLF FFVYENAFLP FTLGIMAIAA
10		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....  3785 3795 3805 3815 3825 3835
15	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	VLTMLKHKF LFLQVFLLEPT VIATALYN-- --CVLDYXIV KFLADHEN-Y NVSVLQMDVQ CLTFVVKHKV LFLQVFLLEPS IIVAAIQN-- --CAWDYHVT KVLAEKFD-Y NVSVLQMDIQ LLMFTLKHKT LFFQVFLIPA LIVTSCIN-- --LAFDVEY NYLAEHFD-Y HVSIMGFNAQ VEVSGIKHKM LFFMSFVLPS VILVTAHN-- --LEWDFSXY ESLQSIVENT NTMFLPVDNQ LAMLVVKHKH LYLTMYITPV LFTLLYNNY- -LVVYKHTFR GYVYAWLSY VPSVEYTYTD LAMLVVKHKH LYLTMYIIPV LFTLLYNNY- -LVVYKHTFR GYVYAWLSY VPSVEYTYTD FAMLLVVKHKH LYLTMEIMPV LCTLEYTNY- -LVVYKQSF GLAYAWLSH VPAVDYTYMD FISFTVVKHM AYMDTFLLEPT LITVIIGVCA EVPIFYNTLI SQVVFILSQW YDPVVDFTMV CAMLVVKHKH AFLCLFLLPS LATVAYFN-- -MVYMPASW MRIMTWLELA DTSLSGYRLK
20		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....  3845 3855 3865 3875 3885 3895
25	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	GLVNLVLCF VVFLH---TW RFSKERETHW FTYVCSLIAV AYTYFYSGD- -----F GFVNIFICLF VALLH---TW RFAKERCTHW CTYLFSLIAV LYTALYSYD- -----Y GLVNIEVCFV VTILHGTYTW REFN-TPASS VTYVALLTA AYNFYASD- -----I GVMLTVFCFI VFTYSVRFF TCKQSWFSLA VTTILVIFNM VKIFGTSDEP WTENQIAFCF EVIYGMILLV GMVFVTLRSI NHDLSFIMF VGRLLSVFSL WYKGSNLEEE -----I EVIYGMILLI GMVFVTLRSI NHDLSFIMF VGRLLSVFSL WYKGSNLEEE -----I EVLVGVLLV AMVFVTLRSI NHDVFSVMFL VGRLLSVLVM WYFGANLEEE -----V PWFMLPLVLY TAPKCVQGCY MNSFNTSLLM LYQFVKLGEV IYTSNTLTA YTEGNWELFF DCVMYASALV LLILMTARTV YDDAARRVWT LMNVITLVYK VYGNALDQA -----I
30		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....  3905 3915 3925 3935 3945 3955
35	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	LSLLVMFLCA ISSDWYIGAI VFLSRLLIF FSPE--SVFS VFGDVKTLV VYLICGYLVC VSLVMFLCA ISNEWYIGAI IFRICRFGVA FLPV--EYVS YFDGVKTVLL FYMLLGFVSC LSCAMTLFAS VTGNWVFGAV CYKAVAYMAL REF---TFVA IFGDIKVMF CYLVLGYFTC VNMLTMIIVSL TTKDWMVIV SYRIAYYIV CUMP-SAFVS DFGFMKCSI VYMAGCYLFC LLMLASLFGT YTWT---TVL SMAVAKVIK WVAVNVLVYF DIPQIKIVLL CYLFIGYIIS LLMLASLFGT YTWT---TAL SMAAAKVIK WVAVNVLVYF DIPQIKIVL CYLFIGYIIS LLFLTSLEGT YTWT---TML SLATAKVIK WLAVNVLVYF DVPQVKLVLL SYLCIGYVCC ELVHTTVLAN VSSNSLIGLF VFCAKWMML YCN-----AT YLNNYVLMV MVNCIGWLCT SMWALVISVT SNYSGVVTI MFLARAIVFV CVEYPLLF TGNLTQICML VYCFGLGYCCC
40		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....  3965 3975 3985 3995 4005 4015
45	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	TYGILYWFN RFFKCTMGVY DFKVSAAEFK YMVANGLHAP YGPFDAWLS FKLLGIGGDR MYYGILYWIN RFCKCTLGVI DFCVSPAEFK YMVANGLNAP NGPFDAWLS FKLLGIGGDR CYGILYWFN RFFKVSUGVY DYTVSAAEFK YMVANGLRAP TGTLDLSSLL AKLIGIGGER CYGILYWFN RFTCMTCGVY QFTVSAAELK YMTANNLSAP KNAYDAMILS AKLIGVGGKR CYWGLFSLMN SLFRMPLGVY NYKISVQELR YMNANGLRPP KNSFEALMLN FKLLGIGGVP CYWGLFSLMN SLFRMPLGVY NYKISVQELR YMNANGLRPP KNSFEALMLN FKLLGIGGVP CYWGLVSLLN SIFRMPLGVY NYKISVQELR YMNANGLRPP RNSFEALVLN FKLLGIGGVP CYFGLYWWVN KVFGTLGKY NFKVSVDQYR YMCLHKNIPP KTVWVVFSTN ILIQGIGGDR CYFGLFCLLN RYFRLTLGVY DYLVSTQEFR YMNQGLLPP KSSIDAFKLN IKLLGIGGKP
50		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....  4025 4035 4045 4055 4065 4075
55	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	CIKISTVQSK LTDLKCTNVV LGLCLSSMNI AANSSEWAYC VDLHNKINLC DDPEKAQGM TIKISTVQSK LTDLKCTNVV LMGILSNMNI ASNSKEWAYC VEMHNKINLC DDPEKAQEL NIKISTVQSK LTDIKCSNVV LGLCLSSMNV SANSTEWAYC VDLHNKINLC DDPEKAQEL NIKISTVQSK LTEMKCTNVV LGLLSKMHV ESNSKEWNYC VGLHNEINLC DDPEIVLEKL IEVVSQFQSK LTDVKCANVV LNCLOHLHV ASNSKLWHYC STLHNEILAT SDLSVAFEKL IEVVSQFQSK LTDVKCANGV LNCLOHLHV ASNSKLWHYC STLHNEILAT SDLSVAFEKL IEVVSQFQSK LTDVKCNVV LNCLOHLHI ASSSKLWQYC STLHNEILAT SDLSVAFDKL VLEIATVQAK LSDVKCTTVV LMQLLTKLV EANSKMHVYL VELHNKILAS DDVGECMDNL CIKVATVQSK MSDVKCTSVV LLSVLQQLRV ESSSKLWQYC VQLHNDILLA KDTTEAFERK
60		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....  4085 4095 4105 4115 4125 4135
65	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	LALLAFFLSK HSDFG----- -LDGLIDSYF DNSSTLQSV SSVSMPSYI AYENARQAYE LALLAFFLSK HSDFG----- -LDGLVDSYF ENDSILQSV SSVSMPSYI AYETARQAYE LALLAFFLSK NSAFG----- -LDDLESYF NDNSMLQSV STYVGLPSY IYENARQAYE LALLAFFLSK HNTCD----- -LSEILSYF ENTILQSV SAYAALPSWI ALEKARADLE AQLLIVLFAN PAAVDSKCLT SIEEVCDDYA KDNTVLQALQ SEFVNMAFV EYEVAKKNLD AQLLIVLFAN PAAVDSKCLT SIEEVCDDYA KDNTVLQALQ SEFVNMAFV EYEVAKKNLD AQLLVVLFAN PAAVDSKCLA SIEEVSDDYV RDSTVLQALQ SEFVNMAFV EYELAKKNLD LGMILTLCFI DSTID----- -LSEYCDIL KRSTVLQSVT QEFSHIPSYA EYERAKKNLYE
70		
75	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV	
80		

69/87

SARS CoV		VSLLSVLLSM QGAVD-----INRLCEEML DNRATLQAI A SEFSSLSPSYA AYATAQEAYE					
5	EMCR	4145	4155	4165	4175	4185	4195
	229E	DAIANGSS--	--SQLIKQLK	RAMNIAKSEF	DHEISVQKKI	NRMAEQAAATQ	MYKEARSVNR
	PEDV	NAVANGSS--	--PQIIKQLK	KAMNVAKAEF	DRESSVQKKI	NRMAEQAAAA	MYKEARAVNR
	TGEV	DAVNGSP--	--PQLVKQLR	HAMNVAKSEF	DREASTQKRL	DRMAEQAAAQ	MYKEARAVNR
	OV43	EAKKNDVS--	--PQILKQLT	KAFNIAKSDF	EREASVQKKL	DKMAEQAAAS	MYKEARAVDR
10	BoCoV	EARFSGSAN--	--QQQLKQLE	KACNIAKSAY	ERDRAVAKKL	ERMADLALTN	MYKEARINDK
	MHV	EACSSGSAN--	--QQQLKQLE	KACNIAKSAY	ERDRAVAKKL	ERMADLALTN	MYKEARINDK
	AIBV	EAKASGSAN--	--QQQIKQLE	KACNIAKSAY	ERDRAVAKKL	ERMADLALTN	MYKEARINDK
	SARS CoV	KVLVDSKNGG	VTQOELAYR	KAANIAKSVF	DRDLAVQKKL	DSMAERAMTT	MYKEARVTDOR
		QAVANGDS--	--EVVLKKLK	KSLNVAKSEF	DRDAAMQKRL	EKMADQAMTQ	MYKQARSEDK
15	EMCR	4205	4215	4225	4235	4245	4255
	229E	KSKVISAMHS	LLFGMLRRLD	MSSVETVLNL	ARDGVVPLSV	IPATSASKLT	IVSPDLSEYS
	PEDV	KSKVVSAMHS	LLFGMLRRLD	MSSVDTILNM	ARNGVVPLSV	IPATSAARLV	VVVPDHDSEFV
	TGEV	KSKIVSAMHS	LLFGMLRRLD	MSSVDTILNL	AKDGVVPLSV	IPAVSATKLN	IVTSDIDSYN
	OV43	KSKVVSALQT	MLFSMVRLKD	NOALNSILDN	AVKGCVPPLNA	IPASATRLV	VITPSLEFVS
20	BoCoV	KSKVVSALQT	MLFSMVRLKD	NOALNSILDN	AVKGCVPPLNA	IPSLAANTLN	IIVPDKSVYD
	MHV	KSKVVSALQT	MLFSMIRKLD	NOALNSILDN	AVKGCVPPLNA	IPSLAANTLT	IIVPDKSVYD
	AIBV	RAKLVSSLHA	LLFSMLKKID	SEKLVNLFEDQ	ASSGVVPLAT	IPSLTSNTLT	IIVPDKQVFD
	SARS CoV	RAKVTSAMQT	MLFTMLRRLD	NDALNNIINN	ARDGCVPLNI	VPIVCSNKLT	LVPDPETWV
						IPLTTAAKLM	VVVPDYGTXY
30	EMCR	4265	4275	4285	4295	4305	4315
	229E	KIVCDGSVHY	AGVVWTLNDV	KDNDGRPVHV	KEITRENVET	LT-----WPL	IINCERVVK-
	PEDV	KMMVDGFFVHY	AGVVWTLQEV	KDNDGKNVHL	KDVTKENQEI	LV-----WPL	ILTCERVVK-
	TGEV	RIOREGCVHY	AGTIWNLIDI	KDNDGKVVHV	KEVTAQNAES	LS-----WPL	VLGGERIVK-
	OV43	KIROENNVHY	AGAIWTIVEV	KDANGSHVHL	KEVTAANELN	LT-----WPL	SITCERTTK
35	BoCoV	QVVDNVYVTY	AGNVWQIQTI	QDSGDTNKQL	NEISDDCN--	-----WPL	VIIANRYNE-
	MHV	QVVDNVYVTY	AGNVWQIQTI	QDSGDTNKQL	HEISDDCN--	-----WPL	VIIANRYNE-
	AIBV	QVVDNVYVTY	AGNVWHIQSI	QDADGAVKQL	NEIDVNLIT--	-----WPL	VIIANRHNE-
	SARS CoV	KCEVEGVHVTY	STVVWNIDTV	IDADGTETLHP	TSTGSGLTTC	ISGANIAWPL	KVNLTNRGNH
		NTCDGNTFTY	ASALWEIQVQ	VDADSKIVQL	SEINMDNSPN	LA-----WPL	IVTALRAN--
40	EMCR	4325	4335	4345	4355	4365	4375
	229E	----LQ-NNE	IMPGKCLKQK	MKAEG--DGG	VLGDGNALYN	TEGGKTFMYA	YISNKADLKF
	PEDV	----LQ-NNE	IMPGKMKVKA	TKGEG--DGG	ITSEGNALYN	NEGGRAFMYA	YVTTKPGMKY
	TGEV	----LQ-NNE	IIPGKCLKORS	IKAEG--DG-	IVGEGKALYN	NEGGRTFMYA	FISDKPDLRV
	OV43	VSATVLQNN	LMPAKLKIQV	VNSGPDQTCN	TPT--QCYYN	NSNNGKIVYA	ILSDVDGLKY
45	BoCoV	VSATVLQNN	LMPAKLKIQV	VNSGPDQTCN	TPT--QCYYN	NSNNGKIVYA	ILSDVDGLKY
	MHV	VSSVVLQNN	LMPQKLRTQV	VNSGSDMNCN	TPT--QCYYN	NSYNGKIVYA	ILSDVDGLKY
	AIBV	KVDVVLQNN	LMPHGVKTKA	CVAGVDQAH	SVES-KCYT	TTGMGKIVYA	ILSDVDGLKY
	SARS CoV	-SAVKLQNN	LSPVALRQMS	CAAGTTQTAC	TDDNALAYN	NSKGGFRVLA	LLSDHQDLKW
50	EMCR	4385	4395	4405	4415	4425	4435
	229E	VKWEYEGG--	CNTIELDSPC	RFMVETPNGP	QVKYLYFVK	LNTLRGAVL	GFIGATIRLQ
	PEDV	VKWEHDSG--	VVTVELEPPC	RFVIDTPTGP	QIKYLYFVK	LNNLRGAVL	GYIGATVRLQ
	TGEV	VKWEFDGG--	CNTIELEPPR	KFLVDSPNGA	QIKYLYFVRN	LNTLRGAVL	GYIGATVRLQ
	OV43	VKWESENND--	IIPIELEAPL	RFYVDGANGP	EVKLYLYFVK	LNTLRGAVL	GYIGATVRLQ
55	BoCoV	TKILKDDGN-	FVLELDPPC	KFTVQDAKGL	KIKYLYFVK	CNTLARGWV	GTISSTVRLQ
	MHV	TKILKDDGN-	FVLELDPPC	KFTVQDAKGL	KIKYLYFVK	CNTLARGWV	GTISSTVRLQ
	AIBV	TKIVKEDGN-	CVLELDPPC	KFSVQDVKGL	KIKYLYFVK	CNTLARGWV	GTISSTVRLQ
	SARS CoV	ASFLNEAGN-	QIYVDLPPC	KFGMKVGK	EVVLYFIK	TRSIVRGMVL	GAISNVVVLQ
		AREPKSDGTG	TIYTELEPPC	REVTDTPKGP	KVKYLYFIK	LNNLRGMVL	GSLAATVRLQ
60	EMCR	4445	4455	4465	4475	4485	4495
	229E	AG-KQTElav	NSGLLTACAF	SVDPAATYLE	AVKHGAKPVS	NCIKMLNSGA	GNGQAITTSV
	PEDV	AG-KQTEFVS	NSHLLTHCSF	AVDPAAAYLD	AVKQGAKPVG	NCVKMLTNGS	GSGQAITCTI
	TGEV	AG-KQTEQAI	NSSLLTLCAF	AVDPAKTYID	AVKSGHKPVG	NCVKMLANGS	GNGQAVTNGV
	OV43	AG-KPTEHPS	NSSLLTLCAF	SPDPAKAYVD	AVKRGMPVN	NCVKMLNSGA	GNGMAVTNGV
70	BoCoV	AG-TATEYAS	NSSILSLCAF	SVDPKKTYLD	FIQGGGTPIA	NCVKMLCDHA	GTGMAITVKP
	MHV	AG-TATEYAS	NSSILSLCAF	SVDPKKTYLD	FIQGGGTPIA	NCVKMLCDHA	GTGMAITVKP
	AIBV	AG-TATEYAS	NSAIRSLCAF	SVDPKKTYLD	YIQGGAPVT	NCVKMLCDHA	GTGMAITVKP
	SARS CoV	SKGHETEVD	AVGILSLCSF	AVDPADTYCK	YVAAGNQLG	NCVKMLTVHN	GSGFAITSKP
		AG-NATEVPA	NSTVLSFCAF	AVDPAKAYKD	YLASGGQPIT	NCVKMLCTHT	GTGQAITVTP
75	EMCR	4505	4515	4525	4535	4545	4555
	229E	DANTNQDSYG	GASICLYCRA	HVPHP-----S	MDGYCKFKGK	CVQVP-IGCL	DPIRFCLENN
	PEDV	DSNTTQDTYG	GASVCIYCRA	HVAHP-----T	MDGFCQYKKG	WVQVP-IGTN	DPIRFCLENT
	TGEV	EASTNQDSYG	GASVCLYCRA	HVEHP-----S	MDGFCRLKKG	YVQVP-LGTV	DPIRFVLEND
		EANTQQDSYG	GASVCIYCRC	HVEHP-----A	IDGLCRYKKG	FVQIP-TGTQ	DPIRFCIENE

70/87

OV43	DATTSQDSYG	GASVCIYCRA	RVEHP----	D	VDGLCKLRGK	FVQVP-VGIR	DPVSYVLTHD
BoCoV	DATTSQDSYG	GASVCIYCRA	RVEHP----	D	VDGLCKLRGK	FVQVP-VGIR	DPVSYVLTHD
MHV	EATTNQDSYG	GASVCIYCRA	RVEHP----	D	VDGLCKLRGK	FVQVP-LGIR	DPVSYVLTHD
AIBV	SPTPDQDSYG	GASVCLYCRA	HIAHPGSVGN		LDGRCCQFKGS	FVQIP-TTEK	DPVGFCLRNK
5 SARS CoV	EANMDQESFG	GASCCLYCRC	HIDHP----	N	PKGFCDLKGK	YVQIPTTCAN	DPVGFTLRNT

	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	4565	4575	4585	4595	4605	4615	
10 EMCR	VCNVCGCWL	HGCACDRTTI	QSVDIS----	YLNRRARGSSA	-ARLEPCN-G	TDIDKCVRAF	
229E	VCKVCGCWL	HGCTCDRTAI	QSFDS----	YLNRRVRGSSA	-ARLEPCN-G	TDIDYCVRAF	
PEDV	VCKVCGCWL	NGCTCDRSIM	QSTDYG----	LFKRVRGSSA	-ARLEPCN-G	TDQHVYRAF	
TGEV	VCVVCWCWL	NGCMCDRTSM	QSFTVDQSY-	LFKRVRGSSA	-ARLEPCN-G	TDPDHVSRAF	
OV43	VCRVCGFWRD	GSCSCVSTDT	TVQSKDTN--	FFKRVRGTSV	DARLVPCASG	LSTDVQLRAF	
BoCoV	VCRVCGFWRD	GSCSCVSTDT	TVQSKDTN--	FFKRVRGTSV	DARLVPCASG	LSTDVQLRAF	
15 MHV	VCRVCGFWRD	GMFLCH-HRL	BVSVKRHE--	LFKRVRGTSV	NARLVPCASG	LSTDVQLRAF	
AIBV	VCTVCGCWIG	YGCCQDSLRL	PKSSVQSG--	----VAGASD	FDKNYLN--	YGVAVRLGMF	
5 SARS CoV	VCTVCGMWKG	YGSCDQLRE	FLMQSADAST	FLNRVCVSA	-ARLTPCGTG	TSTDVVYRAF	

	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	4625	4635	4645	4655	4665	4675	
20 EMCR	DIYNKNVSFL	GKCLKMNVCVR	FKNADLK---	DGYFVIKRC	TKSVMEHEQS	MYNLLNFSGA	
229E	DVYNKDSFI	GKLNKSNVCVR	FKNVDKD---	DAFYIVKRC	IKSVMDEHQS	MYNLLKGCNA	
PEDV	DIYNKDVACL	GKFLKVNVCVR	LKNLDKH---	DAFYIVKRC	TKSAMEHEQS	IYRLEKCGA	
TGEV	DIYNKDVACI	GKFLKTNCSR	FRNLDKH---	DAYYIVKRC	TKTVMDEHQS	CYNDLKDSGA	
25 OV43	DIYNASVAGI	GLHLKVNCCR	FORVDENGDK	LDQFFVVKRT	DLTIYNREMK	CYERVKDCKF	
BoCoV	DICNASVAGI	GLHLKVNCCR	FORVDENGDK	LDQFFVVKRT	DLTIYNREME	CYERVKDCKF	
MHV	DICNANRAGI	GLYKVNCCR	FORADEGNT	LDKFFVIKRT	NLEVYNKEKE	CYELTKECGV	
AIBV	QNLKRNCARF	QELADTEDGN	LEYLDS----	YFVVKQT	TPSNYEHEKS	CYEDLKS-EV	
5 SARS CoV	DIYNEKVAGF	AKFLKTNCCR	FQEKDEEGNL	LDSYFVVKRH	TMSNYQHEET	IYNLVKDCPA	

	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	4685	4695	4705	4715	4725	4735	
30 EMCR	LAEHDFFTWK	DGRVIYGNVS	RHNLTKYTMM	DLVYAMRNF	EQNCDVLKEV	LVLTGCCDNS	
229E	VAKHDFFTWH	EGRTIYGNVS	RQDLTKYTMM	DLCEALRNFD	EKDCVEFKEI	LVLTGCCSTD	
PEDV	IAEHDFFTWK	DGRAIYGNVC	RKDLTEYTMM	DLCYALRNFD	ENNCVVLKSI	LIKVGACEES	
TGEV	VAEHDFFTYK	EGRCFEGNVA	RRLTKYTMM	DLCYAIRNFD	EKNCEVLKEI	LVTVGACTEE	
35 OV43	VAEHDFFTFD	VEGSRVPHIV	RKDLTKYTML	DLCYALRHFD	RNDCMLLCDI	LSIYAGCEQS	
BoCoV	VAEHDFFTFD	VEGSRVPHIV	RKDLTKYTML	DLCYALRHFD	RNDCMLLCDI	LSIYAGCEQS	
MHV	VAEHDFFTFD	VEGSRVPHIV	RKDLTKYTML	DLCYALRHFD	RNDCSTLKEI	LLTYAECDQS	
40 AIBV	TADHDFEFVN	KN---LYNIS	RQRLTKYTMM	DFCYALRHFD	PKDCEVLKEI	LVTYGCIEDY	
5 SARS CoV	VAHDFEFKFR	VDGDMVPHIS	RQRLTKYTMA	DLVYALRHFD	EGNCDTLKEI	LVTYNCCDD	

	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	4745	4755	4765	4775	4785	4795	
45 EMCR	----YFDSKG	WYDPVENEDI	HRVYASLGKI	VARAMLKCV	LCDAMVAKGV	VGVLTLNOD	
229E	----YFEMKN	WEDPIENEDI	HRVYAALGKV	VANAMLKCV	FCDEMVLKGV	VGVLTLNOD	
PEDV	----YFNKNV	WEDPIENEDI	HRVYALGTI	VARAMLKCV	FCDAMVEQGI	VGVLTLNOD	
TGEV	----YFNKND	WEDPIENEDI	HEVYAKLGPI	VANAMLKCV	FCDAIVEKGI	IGVITLNDQ	
50 OV43	----YFTKKD	WYDFVENPDI	INVYKLGPI	FNRLVSATE	FADKLVEVGL	VGVLTLNOD	
BoCoV	----YFTKKD	WYDFVENPDI	INVYKLGPI	FNRLVSATE	FADKLVEVGL	VGVLTLNOD	
MHV	----YFQKKD	WYDFVENSDI	INVYKLGPI	FNRLVNTAK	FADTLVEAGL	VGVLTLNOD	
AIBV	HPKWFEENKD	WYDFVENSKY	YVLMKMGPI	VRRALLNAIE	FGNLMVEKGI	VGVLTLNOD	
5 SARS CoV	----YFNKKD	WYDFVENPDI	LRVYANLGER	VRQSLKTVQ	FCDAMRDAGI	VGVLTLNOD	

	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	4805	4815	4825	4835	4845	4855	
55 EMCR	LNGNFDYDFD	FVSLPNMGV	PCCTSYYSYM	MPIMGLTNCL	ASECFVKS	FGSDFKTFDL	
229E	LNGNFDYDFD	FVLCPPGMGI	PYCTSYYSYM	MPVMGMTNCL	ASECFMKS	FGQDFKTFDL	
PEDV	LNGDFFYDFD	FTCSIKGMV	PICTSYYSYM	MPVMGMTNCL	ASECFVKS	FGEDFKSYDL	
TGEV	LNGNFDYDFD	FVKTAPGFC	ACVTSYYSYM	MPLMGMTNCL	ESENFKVSDI	YGSYKQYDL	
60 OV43	LNGKWDYDFD	YVIAAPGCGV	AIADSYYSYM	MPMLTMCHAL	DCELYVNN--	---AYRLFDL	
BoCoV	LNGKWDYDFD	YVIAAPGCGV	AIADSYYSYM	MPMLTMCHAL	DCELYVNN--	---AYRLFDL	
MHV	LYGQWYDFD	FVKTVPGCGV	AVADSYYSYM	MPMLTMCHAL	DSELFING--	---TYREFDL	
AIBV	LNGKFDYDFD	FQKTAPGAGV	PVFTYYSYM	MPILAMTDAL	APERYFEYDV	H-KGYKSYDL	
65 SARS CoV	LNGNWDYDFD	FVQVAPGCGV	PIVDSYYSYL	MPILTLTRAL	AAESHMDADL	A-KPLIKWDL	

	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	4865	4875	4885	4895	4905	4915	
70 EMCR	LKYDFTEHKE	NLFNKYFKHW	SFDYHPNCSD	CYDDMCVHC	ANFNTLFATT	IPGTAFGPLC	
229E	LKYDFTEHKE	NLFNKYFKYW	QDYHPDCVD	CHDEMCLHC	SNFNTLFATT	IPNTAFGPLC	
PEDV	LEYDFTEHKT	ALFNKYFKYW	GLQYHPNCVD	CSDEQCIVHC	ANFNTLFSTT	IPITAFGPLC	
TGEV	LAYDFTEHKE	YLFQKYFKYW	DRTYHPNCSD	CTSDCCIHC	ANFNTLFSMT	IPMTAFGPLV	
75 OV43	VQYDFTDYKL	ELFNKYFKHW	SMPYHPNTVD	CQDDRCIHC	ANFNILFSMV	LPNTCFGPLV	
BoCoV	VQYDFTDYKL	ELFNKYFKHW	SMPYHPNTVD	CQDDRCIHC	ANFNILFSMV	LPNTCFGPLV	
MHV	VQYDFTDYKL	ELFNKYFKYW	SMTYHPNTCE	CEDDRCIHC	ANFNILFSMV	LPNTCFGPLV	
AIBV	LKYDYTEEKQ	ELFNKYFKYW	DQYHPNCRD	CSDDRCIHC	ANFNILFSTL	IPQTSFGLNLC	
5 SARS CoV	LKYDFTEERL	CLFDRYFKYW	DQTYHPNCIN	CLDDRCIHC	ANFNVLFTSV	FPPTSFGPLV	

	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	4925	4935	4945	4955	4965	4975	





72/87

SARS CoV		NSVFNICQAV TANVNALLST DGNKIADKYV RNLQHRLYEC LYRNRDVE DHE FVDEFYAYLR					
5	EMCR	5345	5355	5365	5375	5385	5395
	229E	KHFSMMILSD	DGVVCYNKDY	AELGYIADIS	AFKATLYYQN	NVFMSTSKCW	VEEDLTGKGP
	PEDV	KHFSMMILSD	DSVVCYNKTY	AGLGYIADIS	AFKATLYYQN	GVFMSTAKCW	TEEDLSIGPH
	TGEV	KHFSMMILSD	DGVVCYNNDY	ASLGYYADLN	AFKAVLYYQN	NVFMASAKCW	IEPDINKGPH
	OV43	KHFSMMILSD	DGVVCYNSDY	ASKGYIANIS	AFQOVLYYQN	NVFMSTSKCW	VEPDLVSGPH
10	BoCoV	KHFSMMILSD	DGVVCYNSDY	ASKGYIANIS	AFQOVLYYQN	NVFMSESKCW	VEHDINNGPH
	MHV	KHFSMMILSD	DGVVCYNSEF	ASKGYIANIS	AFQOVLYYQN	NVFMSEAKCW	VETDIEKGP
	AIBV	KNFSLMILSD	DGVVCYNNLT	AKQGLVADIS	GFREVLYYQN	NVFMADSKCW	VEPDLEKGP
	SARS CoV	KHFSMMILSD	DAVVCYNSNY	AAQGLVASIK	NFKAVLYYQN	NVFMSEAKCW	TETDLTKGPH
		5405	5415	5425	5435	5445	5455
20	EMCR	EFCSQHTMQI	VDKDGTYYLP	YPDPSRILSA	GVFVDDVVK	DAVVLXRYV	SLAIDAYPLS
	229E	EFCSQHTMQI	VDENGKYYLP	YPDPSRIISA	GVFVDDITKT	DAVILLERYV	SLAIDAYPLS
	PEDV	EFCSQHTMQI	VDKEGTYYLP	YPDPSRILSA	GVFVDDVVK	DAVLLERYV	SLAIDAYPLS
	TGEV	EFCSQHTLQI	VGPDDGYLLP	YPDPSRILSA	GVFVDDIVKT	DNVIMLERYV	SLAIDAYPLT
	OV43	EFCSQHTMLV	KMDGDDVYLP	YPNPSRILGA	GCFVDDLKT	DSVLLIERFV	SLAIDAYPLV
25	BoCoV	EFCSQHTMLV	KMDGDDVYLP	YPVPSRILGA	GCFVDDLKT	DSVLLIERFV	SLAIDAYPLV
	MHV	EFCSQHTMLV	KMDGDEVYLP	YPDPSRILGA	GCFVDDLKT	DSVLLIERFV	SLAIDAYPLV
	AIBV	EFCSQHTMLV	EVDGEPKYL	YPDPSRILGA	GVFVDDVKT	EPVAVMERYI	ALADAYPLV
	SARS CoV	EFCSQHTMLV	KQGDYVYLP	YPDPSRILGA	GCFVDDIVKT	DGTLMIERFV	SLAIDAYPLT
		5465	5475	5485	5495	5505	5515
30	EMCR	KHENSEYRKV	FYVLLDWVKH	LNKNLNEGVL	ESFSVTLLDN	QEDKFWCEDF	YASMYENSTI
	229E	KHPKPEYRKV	FYALLDWVKH	LNKTLNEGVL	ESFSVTLLDE	HESKFWDESF	YASMYEKSTV
	PEDV	KHENPEYKKV	FYVLLDWVKH	LYKTLNAGVL	ESFSVTLLDE	STAKFWDESF	YANMYEKSAV
	TGEV	KHKPAYQKV	FYTLDDWVKH	LQNLNAGVL	DSFSVTMLLE	GQDKFWSEEF	YASLYEKSTV
	OV43	YHENEYQKV	FRVYLAYIKK	LYNDLGNQIL	DSYSVILSTC	DGQKFTDESF	YKNMYLRSV
35	BoCoV	YHENEYQKV	FRVYLEYIKK	LYNDLGNQIL	DSYSVILSTC	DGQKFTDESF	YKNMYLRSV
	MHV	YHENPEYQNV	FRVYLEYIKK	LYNDLGNQIL	DSYSVILSTC	DGQKFTDET	YKNMYLRSV
	AIBV	HHENEYKKV	FFVLLAYIRK	LYQELSONML	MDYSFVMDID	KGSKFWEQEF	YENMYRAPTT
	SARS CoV	KHPNQEYADV	FHLYLQYIRK	LHDELTGHML	DMYSVMLTND	NTSRYWEPEF	YEAMYTPHTV
		5525	5535	5545	5555	5565	5575
40	EMCR	LQAAGLCVVC	GSQTVLRCGD	CLRKPMCTK	CAYDHVFGTD	HKFILAITYP	VCNAGSGCVS
	229E	LQAAGLCVVC	GSQTVLRCGD	CLRRPMLCTK	CAYDHVFGTD	HKFILAITYP	VCNTSGCNVN
	PEDV	LQAAGLCVVC	GSQTVLRCGD	CLRRPMLCTK	CAYDHVIGTT	HKFILAITYP	VCNASCDCVN
	TGEV	LQAAGMCVVC	GSQTVLRCGD	CLRRPMLCTK	CAYDHVMGKT	HKFIMSITPY	VCSFNGCNVN
	OV43	MQSVGACVVC	SSQTSILRCGS	CIRKPLCCK	CCYDHVMATD	HKYVLSVSPY	VCNAPGCDVN
45	BoCoV	MQSVGACVVC	SSQTSILRCGS	CIRKPLCCK	CCYDHVMATD	HKYVLSVSPY	VCNAPGCDVN
	MHV	MQSVGACVVC	SSQTSILRCGS	CIRKPLCCK	CAYDHVMSTD	HKYVLSVSPY	VCNSPGCDVN
	AIBV	LQSCGVCVVC	NSQTLIRCGN	CIRKPLCCK	CCYDHVMHTD	HKNVLSINPY	ICSQGLGCEA
	SARS CoV	LQAVGACVLC	NSQTSILRCGA	CIRRPFLCCK	CCYDHVISTS	HKLVLVSNPY	VCNAPGCDVT
		5585	5595	5605	5615	5625	5635
55	EMCR	DVTKLYLGGL	NYCYTNHKKPQ	LSFPLCSAGN	IFGLYKNSAT	GSLDVEVFNR	LATSDWTDVR
	229E	DVTKLYLGGL	NYCYVDHKKP	LSFPLCSAGN	VFGLYKSSAL	GSMDDIDVFNK	LSTDWSDIR
	PEDV	DVTKLYLGGL	SYWCHCHKPR	LAFPLCSAGN	VFGLYKNSAT	GSPDVEDEFNR	IATSDWTDVS
	TGEV	DVTKFLGGL	SYCYMNHKKPQ	LSFPLCANGN	VFGLYKSSAV	GSEAVEDEFNK	LAVSDWTNVE
	OV43	DVTKLYLGGM	SYCEDHKKPQ	YSFKLVNMGL	VFGLYKQSC	GSPYIDDEFNR	IASCWTDVD
60	BoCoV	DVTKLYLGGM	SYCEDHKKPQ	YSFKLVNMGM	VFGLYKQSC	GSPYIDDEFNR	IASCWTDVD
	MHV	DVTKLYLGGM	SYCEDHKKPQ	YSFKLVNMGM	VFGLYKQSC	GSPYIDDEFNR	IASCWTEVD
	AIBV	DVTKLYLGGM	SYFCGNHKKP	LSIPLVSNGT	VFGIYRANCA	GSENVDDFNQ	LATTNWSIVE
	SARS CoV	DVTQLYLGGM	SYCKSHKPP	ISFPLCANGN	VFGLYKNTCV	GSDNVTDFNA	IATCDWTNAG
		5645	5655	5665	5675	5685	5695
65	EMCR	DYKLANDVKD	TLRLFAAETI	KAKEESVKSS	YAFATLKEIV	GPKELLLSWE	SGKVKPPLNR
	229E	DYKLANDAKE	SLRLFAAETV	KAKEESVKSS	YAYATLKEIV	GPKELLLSWE	SGKAKPPLNR
	PEDV	DYRLANDVKD	SLRLFAAETI	KAKEESVKSS	YACATLHEVV	GPKELLLSWE	VGRPKPPLNR
	TGEV	DYKLANNVKE	SLKIFAETV	KAKEESVKSE	YAYAVLKEVI	GPKELVLSWE	ASKTKPPLNR
	OV43	DYILANECTE	RLKLFAAETQ	KATEEAFKQS	YASATIQEV	SERELILSWE	IGKVKPPLNK
70	BoCoV	DYILANECTE	RLKLFAAETQ	KATEEAFKQS	YASATIQEV	SERELILSWE	IGKVKPPLNK
	MHV	DYVLANECTE	RLKLFAAETQ	KATEEAFKQS	YASATIREIV	SDRELILSWE	IGKVRPPLNK
	AIBV	PYILANRCS	SLRRFAAETV	KATEELHKQ	FASAEVREVF	SDRELILSWE	PGKTRPPLNR
	SARS CoV	DYILANTCTE	RLKLFAAETL	KATEETFKLS	YGIATVREVL	SDRELILSWE	VGKPRPPLNR
		5705	5715	5725	5735	5745	5755
75	EMCR	NSVFTCFQIS	KDSKFQIGEF	IFEKVEYGS	TVTYKSTVTT	KLVPGMIFVL	TSHNVQPLRA
	229E	NSVFTCFQIT	KDSKFQVGEF	VFEKVDYGS	TVTYKSTATT	KLVPGMIFVL	TSHNVQPLRA
	PEDV	NSVFTCYHIT	KNTKFQIGEF	VFEKAEYDND	AVTYKTATT	KLVPGMIFVL	TSHNVQPLRA
	TGEV	NSVFTCFQIS	KDTKIQLGEF	VFEQSEYGS	SVYYKSTSTY	KLTPGMIFVL	TSHNVSPKLA
		5705	5715	5725	5735	5745	5755
80	EMCR	NSVFTCFQIS	KDSKFQIGEF	IFEKVEYGS	TVTYKSTVTT	KLVPGMIFVL	TSHNVQPLRA
	229E	NSVFTCFQIT	KDSKFQVGEF	VFEKVDYGS	TVTYKSTATT	KLVPGMIFVL	TSHNVQPLRA
	PEDV	NSVFTCYHIT	KNTKFQIGEF	VFEKAEYDND	AVTYKTATT	KLVPGMIFVL	TSHNVQPLRA
	TGEV	NSVFTCFQIS	KDTKIQLGEF	VFEQSEYGS	SVYYKSTSTY	KLTPGMIFVL	TSHNVSPKLA
		5705	5715	5725	5735	5745	5755





74/87

5	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	-----DLHSS -QVCGLFKNC TRTPLNLPT HAHTFLSLSD QFKTTGDLAV QIGS-N-NVC -----DLQSE -SSCGLFKDC ARNPIDLPPS HATTYLSLSD REKTSGLDVA QIGN-N-NVC -----DLQAN -EGCGLFKDC SRGDDLLPPS HANTFMSLAD NEKTDQYLAV QIGV-N-GPI KI---GLQAK PETCGLFKDC SKSEQYIPPA YATTYMSLSD NEKTSGLAV NIG--T-KDV KVPQAVETKV QCSTNLFKDC SKSYSGYHPA HAPSFLAVDD KYKATGDLAV CLGIGD-SAV KVPQAVETRV QCSTNLFKDC SKSYSGYHPA HAPSFLAVDD KYKATGDLAV CLGIGD-SAV KIN---NPRL QCTTNLFKDC SRSYAGYHPA HAPSFLAVDD KYKVGGLAV CLNVAD-SAV T-----S-- LQGTGLFKIC NKEFSGVHPA YAVTTKALAA TYKVNDLAA LVNVEAGSEI RRN-VATLQA ENVTGLFKDC SKIITGLHPT QAPTHLSVDI KFKTEG-LCV DIPGIP-KDM
10		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....  6185 6195 6205 6215 6225 6235
15	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	TYEHVISFMG FRFDISIPGS HSLFCTRDEA IRNVRGWLGM DVESAHVCGD NIGTNVPLQV TYEHVISYMG FRFDVSMPPS HSLFCTRDEA MRHVRGWLGM DVEGAHVCGD NVGTNVPLQV KYEHVISFMG FRFDINIPNH HTLFCTRDEA MRNVRGWLGF DVEGAHVCGD NVGTNVPLQL KYANVISYMG FRFEANIPGY HTLFCTRDEA MRNVRAWLGF DVEGAHVCGD NVGTNVPLQL TYSRLISLMG FKLDVTLDDGY CKLFITKEEA VKRVRAWVGF DAEGAHAATRD SIGTNFPLQL TYSRLISLMG FKLDVTLDDGY CKLFITKEEA VKRVRAWVGF DAEGAHAATRD SIGTNFPLQL TYSRLISLMG FKLDLTLDDGY CKLFITRDEA IRNVRAWVGF DAEGAHAATRD SIGTNFPLQL TYKHLISLLG FKMSVNVGCG HNMFITRDEA IRNVRGWVGF DVEATHACGT NIGTNLFFQV TYRRLISMMG FKMNYQVNGY PNMFITRDEA IRHVRWVGF DVEGCHATRD AVGTNLPLQL
20		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....  6245 6255 6265 6275 6285 6295
25	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	GFSNGVNFVV QTEGCVSTNF GDVIKPVCAR SPPGEQFRHL VFELRKGQPW LIVRRRIVQM GFSNGVDFVA QPEGCVLTNT GSVVKPVRAR APPGEQFTHI VFLLRKGQPW SVLRKRIVQM GFSNGVDFVV RPEGCVVTES GNDYIKPVRAR APPGEQFAHL VFLLRKGQPW DVVRKRIVQM GFSNGVDFVV QTEGCVITEK GNSIEVVKAR APPGEQFAHL VFLLRKGQPW HVRRRIVQM GFSTGIDFVV EATGLFADRD GYSFKKAVAK APPGEQFKHL VFLLRKGQPW DVVRPRIVQM GFSTGIDFVV EATGLFADRD GYSFKKAVAK APPGEQFKHL VFLLRKGQPW DVVRPRIVQM GFSTGIDFVV EATGMFAERD GYVEKKAVAR APPGEQFKHL VFLLRKGQPW DVVRIRIVQM GFSTGADFVV TPEGLVDTSI GNNFEPVNSK APPGEQFNHL RVLFKSAKPW HVIRPRIVQM GFSTGVNLVA VPTGYVDTEN NTEFTRVNAK PPPGDQFKHL VFLLRKGQPW NVVRIKIVQM
35		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....  6305 6315 6325 6335 6345 6355
40	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	ISDYLNLSD ILVFLWAGS LETTMRIFY KIGP-IKICY CGNSATCYNS VSNEYCCFKH IADFLAGSSD VLVEFLWAGG LETTMRIFY KIGA-VKHQ CGTVATCYNS VSNDYCCFKH CSDYLANLSD ILIFVLWAGG LETTMRIFY KIGP-SKSCD CGKVATCYNS ALHTYCCFKH VCDYFDGLSD ILIFVLWAGG LETTMRIFY KIGR-PQKCE CGKSATCYNS QSQVYACFKH FADHLIDLSD CVVLVTWAAN FELTCLRYFA KVGREISCNV CTKRATVYNS RTGYGICWRH FADHLIDLSD CVVLVTWAAN FELTCLRYFA KVGREISCNV STKRATAYNS RTGYGICWRH LSDHLVDLAD SVVLVTWAAS FELTCLRYFA KVGKEVVCV CNKRATCFNS RTGYGICWRH LADNLNVSD CVVFTWCHG LETTCLRYFA KIGK-EQVCS CGSRATTFNS HTQAYACWKH LSDTLKGLSD RVVFLWAGG FELTSMKYFV KIGPERTCCL CDKRATCFST SSDTYACWNH
45		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....  6365 6375 6385 6395 6405 6415
50	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	ALGCDYVYNP YAFDIQOWGY VGSLSQNHHT FCNIHRNEHD ASGDAVMTRC LAVHDCFVKH ALGCDYVYNP YVIDIQOWGY VGSLSQNHHA ICNVHRNEHV ASGDAIMTRC LAVYDCFVKH ALGCDYLYNP YCIDIQOWGY KGSLSLNHHE HCNVHRNEHV ASGDAIMTRC LAIHDCFVKH ALGCDYLYNP YCIDIQOWGY TGSLSMNHHE VCNHRNEHV ASGDAIMTRC LAIHDCFVKH SVTCDYLYNP LIVDIQOWGY IGSLSNHDH YCSVHKGAVH ASSDAIMTRC LAVYDCFCNN SVTCDYLYNP LIVDIQOWGY IGSLSNHDH YCSVHKGAVH ASSDAIMTRC LAVYDCFCNN SYSCDYLYNP LIVDIQOWGY TGSLSNHDH YCSVHKGAVH ASSDAIMTRC LAVYDCFCNN CLGFDYVYNP LLVDIQOWGY SGNLQFNHDL HCNVHGHAVH ASVDAIMTRC LAINNAFCQD SVGFDYVYNP FMIDVQOWGF TGNLQSNHDH HCQVHGHAVH ASCDAIMTRC LAVHECFVKH
60		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....  6425 6435 6445 6455 6465 6475
65	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV SARS CoV	VDWTVTYPI ANEKFINGCG RNVQGHVVR ALKLYKPSVI HDIGNPKGVR CA-VTDAKWY VDWSITYPMI ANENAINKGG RTVQSHIMRA AIKLYNPKAI HDIGNPKGIR CA-VTDAKWY VDWSITYPFI GNEAVINKSG RIVQSHTMRS VLKLYNPKAI YDIGNPKGIR CA-VTDAKWY VDWSIVYPI DNEEKINKAG RIVQSHVMKA ALKIFNPAAI HDVGNPKGIR CA-TTPIPIWF INWNVEYPII SNELISINTSC RVLQRVILKA AMLCNRYTLC YDIGNPKAIA CV--KDFDFK INWNVEYPII SNELISINTSC RVLQRVMLKA AMLCNRYTLC YDIGNPKAIA CV--KDFDFK VNWSLEYPII SNEVSNTSC RLLQRMVFRA AMLCNRYDVC YDIGNPKGLA CV--KGYDFK VNWDLYPHI ANEDEVNSSC RYLQRMVYNA CVDALKVNVV YDIGNPKGK CVRRGDVNF VDWSVEYPII GDELRVNSAC RKVQHMVVK ALLADKFPVL HDIGNPKAIK CVPQAEVEWK
70		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....  6485 6495 6505 6515 6525 6535
75	EMCR 229E PEDV TGEV OV43 BoCoV MHV AIBV	CYDKQPVNSN ---VKLLDYD YATHG---QLD GLCLFWNCNV DMYPEFSIVC RFDTRTRSVF CYDKNPINSN ---VKLTLEYD YMTHG---QMD GLCLFWNCNV DMYPEFSIVC RFDTRTRSTL CFDKNPNTSN ---VKLTLEYD YITHG---QFD GLCLFWNCNV DMYPEFSIVC RFDTRCRSPL CYDRDPINNN ---VRCLDYD YMVHG---QMN GLMLFWNCNV DMYPEFSIVC RFDTRVLSKL FYDAQPIVKS ---VKTLLEYD FEAHKDSFKD GLCMFWNCNV DKYPANAVVC RFDTRVLNNL FYDAQPIVKS ---VKTLLEYD FEAHKDSFKD GLCMFWNCNV DKYPANAVVC RFDTRVLNNL FYDASPVKKS ---VKQFVYK YEAHKQDFLD GLCMFWNCNV DKYPANAVVC RFDTRVLNNL FYDKNPVIRN ---VKQFEYD YNQHKKDFAD GLCMFWNCNV DCYPDNLSVC RYDTRNLVSF
80		

SARS CoV	FYDAQPCSDK	AYKIEELFYS	YATHHDKFTD	GVCLFWNCNV	DRYPANAIVC	RFDTRVLSNL
	6545	6555	6565	6575	6585	6595
EMCR	NLEGVNGGSL	YVNKHAFHTP	AYDKRAFVKL	KPMPPFFYFDD	SDCDVVQ---	-EQVNVVPLR
229E	NLEGVNGGSL	YVNNHAFHTP	AYDKRAMAKL	KPAPFFYYDD	GSCEVHV---	-DOVNVVPLR
PEDV	NLEGCNGGSL	YVNNHAFHTP	AFDKRAFAKL	KMPFFFFYYDD	TECDKLQ---	-DSINYVPLR
TGEV	SLEGCNGGAL	YVNNHAFHTP	AYDRRAFAKL	KPMPPFFYYDD	SNCELVD---	-GQPNYVPLK
OV43	NLPGCNGGSL	YVNKHAFHTK	PFAARAEHL	KMPFFFFYYSD	TPCVYMDGMD	AKQVDYVPLK
BoCoV	NLPGCNGGSL	YVNKHAFHTK	PFSAARAEHL	KMPFFFFYYSD	TPCVYMDGMD	AKQVDYVPLK
MHV	NLPGCNGGSL	YVNKHAFHTS	PFTRAAPENL	KMPFFFFYYSD	TPCVYMEGME	SKQVDYVPLR
AIBV	NLPGCNGGSL	YVNKHAFYTP	KFDRISFRNL	KAMPFFFYDS	SPCETIQVDG	-VAQDLVSLA
SARS CoV	NLPGCDGGSL	YVNKHAFHTP	AFDKSAFTNL	QQLPFFYYSD	SPCESHGKQV	VSDIDYVPLK
	6605	6615	6625	6635	6645	6655
EMCR	ASSCVTRCNI	GGAVCSKHAN	LYQKYVEAYN	TFTQAGFNW	VPHSFDVYNL	WQIFIFT-NL
229E	ATNCITRCNI	GGAVCSKHAN	LYRAYVESYN	IFTQAGFNW	VPTTFDCYNL	WQTFTEV-NL
PEDV	ASNCITRCNV	GGAVCSKHCA	MYRYSVYAYN	TFTSAGFTIW	VPTSFDTYNL	WQTFSN--NL
TGEV	SNVCITRCNI	GGAVCKKHAA	LYRAYVEDYN	IFMQAGFTIW	CPQNFDTYML	WHGFEVNSKAL
OV43	SATCITRCNL	GGAVCLKHAE	EYREYLESYN	TATTAGFTTFW	VYKTFDFYNL	WNTFTK---L
BoCoV	SATCITRCNL	GGAVCLKHAE	EYREYLESYN	TATTAGFTTFW	VYKTFDFYNL	WNTFTK---L
MHV	SATCITRCNL	GGAVCLKHAE	EYREYLESYN	TATTAGFTTFW	VYKTFDFYNL	WNTFTK---L
AIBV	TKDCITRCNI	GGAVCKKHAQ	MYAEFVTSYN	AAVTAGFTTFW	VTNKNLPYNL	WKSFSFA---L
SARS CoV	SATCITRCNL	GGAVCRHHAN	EYRQYLDAYN	MMISAGFSLW	IYKQEDTYNL	WNTFTK---L
	6665	6675	6685	6695	6705	6715
EMCR	QSLENIAFNV	VKKGCFTGVD	GELPVAVVND	KVFVRYGDVD	NLVEFTNKTTL	PTNVAFELFA
229E	QSLENIAFNV	VNKGSEFVGAD	GELPVAISGD	KVFVRDGNVD	NLVEFVNKTSL	PTNIAFELFA
PEDV	QSLENIAFNV	LKKGSFVGDE	GELPVAVVND	KVLVRDGNVD	TLVFTNKTSL	PTNIAFELFA
TGEV	QSLENIAFNV	VKKGATFGLK	GDLPTAVIAD	KIMVRDGNVD	KCIETNKTSL	PTNIAFELFA
OV43	QSLENVVYNL	VKTGHYTGQA	GEMPCAIIIND	KVVAKIDKED	VVIFINNTTY	PTNVAFELFA
BoCoV	QSLENVVYNL	VKTGHYTGQA	GEMPCAIIIND	KVVAKIDKED	VVIFINNTTY	PTNVAFELFA
MHV	QSLENVVYNL	VNAGHFDGRA	GELPCAIVIGE	KVIAKIQNED	VVVFKNNTPE	PTNVAFELFA
AIBV	QSIDNIAYN	YKGGHYDAIA	GEMPTVITIG	KVFVIDQGVV	VVFEVNTTTL	PTSVAFELFA
SARS CoV	QSLENVAYNV	VNKGHFDGHA	GEAPVSIINN	AVYTKVDGID	VEIFENKTTL	PVNVAFELFA
	6725	6735	6745	6755	6765	6775
EMCR	KRKMGTLTPPL	SILKNLGVVA	TYKFWLWDYE	AERPFTSYTK	SVCKYTDEN	-----EDV
229E	KRKVGTLTPPL	SILKNLGVVA	TYKFWLWDYE	AERPLTSFTK	SVCGYTDFA	-----EDV
PEDV	KRKVGTLTPPI	TILRNLGVVC	TSKCVLWDYE	AERPLTTFTEK	DVCKYTDFA	-----GDV
TGEV	KRKLGLTPPL	TILRNLGVVA	TYKFWLWDYE	AERPFSNFTK	QVCSYTDLD	-----SEV
OV43	KRSVRHHPPEL	KLFERNLNDV	CWKHVIWDYA	RESIFCSNTY	GVCMYTDLK	-----FIDKL
BoCoV	KRSIRHHPPEL	KLFERNLNDV	CWKHVIWDYA	RESIFCSNTY	GVCMYTDLK	-----LIDKL
MHV	KRSIRHHPPEL	KLFERNLNDV	CWKHVIWDYA	RESIFCSNTY	GVCMYTDLK	-----LIDKL
AIBV	KRNIRTLPPN	RILKGLGVDV	TNGFVIWDYA	NQTPLYRNTV	KVCKYTDLQ	-----CIESL
SARS CoV	KRNIKVPPEI	KILNNLGVDI	AANTVIWDYK	REAPAHVSTI	KVCAYTDIE	-----PNGL
	6785	6795	6805	6815	6825	6835
EMCR	CVCFDNSIQG	SYERFTLTNT	AVLFSATVVK	N---LTPIK	LNFGMLNGMP	VSSIKSDKG
229E	CTCYDNSIQG	SYERFTLTNT	AVLFSATAVK	TGGKSLPAIK	LNFGMLNGNA	IATVKSDEGN
PEDV	CTLFDNSIVG	SLEFRSMTQNT	AVLMSLTAVK	K---LTGIK	LTYGVLNGVP	VN---THED
TGEV	VTCFDNSIAG	SFERFTTTRD	AVLISNNNAV	G---LSAIK	QYGLLNDLP	VS---TVGN
OV43	NVLFDGRDNG	ALEAFKRSNN	GVYISTTKVK	S-----LS	MIRGPPRAEL	NGVVVDKVD
BoCoV	NVLFDGRDNG	ALEAFKRSNN	GVYISTTKVK	S-----LS	MIRGPPRAEL	NGVVVDKVD
MHV	NVLFDGRDNG	ALEAFKRCRD	GVYINTTTIK	S-----LS	MIRGPPRAEL	NGVVVDKVD
AIBV	VVLYDDR-YG	DYQSLAAND	AVLSTQCYK	R-----YS	YVEIPSNLLV	QNGMPLKDG
SARS CoV	TVLFDGRVEG	QVDLFRNARN	GVLITEGSVK	G-----LT	PSKGPAQASV	NGVTLLIGES
	6845	6855	6865	6875	6885	6895
EMCR	EKLVNWTYTV	RKNGQFQDHY	DG-----			-----FYTQ
229E	IKNINWFYTV	RKDGRKPDHY	DG-----			-----FYTQ
PEDV	-----KPFYTWIYT	RKNGKFEDYP	DG-----			-----FYTQ
TGEV	-----KPVYTWIYT	RKNGEYVEQI	DS-----			-----YTTQ
OV43	-----TDC					

76/87

OV43	SRVSSFTCR	TDMEKDFIAL	DQDVFIQYK	LEDYAFEHIV	YGNFNQKIIG	GLHLLIGLYR
BoCoV	SRVSSFTCR	TDMEKDFIAL	DQDVFIQYK	LEDYAFEHIV	YGNFNQKIIG	GLHLLIGLYR
MHV	SRFLSFAPR	SEMEKDFMDL	DEDVFIKYS	LQDYAFEHV	YGSFNQKIIG	GLHLLIGLAR
AIBV	GRSYETFEPR	SDIERDFLAM	SEESFVERYG	-KDLGLQHIL	YGEVDKPOLG	GLHTVIGMYR
5 SARS CoV	SRDLEDFKPR	SQMETDFLEL	AMDEFIQRYK	LEGYAFEHIV	YGDFSHGQLG	GLHLMIGLAK
	6965	6975	6985	6995	7005	7015
10 EMCR	LSKMGVLKAD	DEVTASDTTL	RCCTVTYLNE	LSSKVVCITYM	DLLDDDFVTI	LK---SLDLG
229E	LSKMGILKAE	EFVAASDITL	KCCTVTYLND	PSSKTVCITYM	DLLDDDFVSV	LK---SLDLT
PEDV	LACMGVLKID	EFVSSNDSTL	KSCVTYADN	PSSKMVCITYM	DLLDDDFVSI	LK---SLDLS
TGEV	LAKMGLFSVQ	EFMNNSDSTL	KSCCITYADD	PSSKNVCITYM	DILLDDDFVTI	IK---SLDLN
OV43	RQQTSNLVVQ	EFVS-YDSSI	HSYFITDEKS	GGSKSVCTVI	DILLDDDFVAL	VK---SLNLN
BoCoV	RQQTSNLVQ	EFVS-YDSSI	HSYFITDEKS	GGSKSVCTVI	DILLDDDFVAL	VK---SLNLN
15 MHV	RQQKSNLVQ	EFVP-YDSSI	HSYFITDENS	GGSKSVCTVI	DLLDDDFVDI	VK---SLNLN
AIBV	LLRANKLNAK	SVTN-SDSDV	MONYFVLSDN	GSYKQVCTV	DLLDDDFLEL	LRNLIKKEYGT
5 SARS CoV	RSQDSPLKLE	DFIP-MDSTV	KNYFITDAQT	GSSKVCVSVI	DLLDDDFVEI	IK---SQDLS
	7025	7035	7045	7055	7065	7075
20 EMCR	VISKVHEVII	DNKPYRWMLW	CKDNHLSFTY	PQLQS-AEWK	CGYAMPQIYK	LQRMCLPECN
229E	VVSKVHEVII	DNKPYRWMLW	CKDNVATFFY	PQLQS-AEWK	CGYSMPGIYK	TQRMCLPECN
PEDV	VVSKVHEVMV	DKMWRWMLW	CKDHKLQTFY	PQLQA-SEWK	CGYSMPGIYK	TQRMCLPECN
TGEV	VVSKVVDVIV	DKAWRWMLW	CENSHIKTFY	PQLQS-AEWN	PGYSMPPLYK	IQRMCLERCN
25 OV43	CVSKVNVNV	DFKDFQFMLW	CNDEKVMTFY	PRLQAASDWK	PGYSMPVLYK	YLNSPMERVS
BoCoV	CVSKVNVNV	DFKDFQFMLW	CNDEKVMTFY	PRLQAASDWK	PGYSMPVLYK	YLNSPMERVS
MHV	CVSKVNVNV	DFKDFQFMLW	CNEEKVMTFY	PRLQAASDWK	PGYVMPVLYK	YLNSPMERVS
AIBV	NKSKVTVTSI	DYHSINEMTW	FEDGSIKTCY	PQLQS-AWT	CGYNMPELYK	VQNCVMEPCN
5 SARS CoV	VISKVVKVTI	DYAEISFMLW	CKDGHVETFY	PKLQASQAWQ	PGVAMPNLYK	MQRMILLEKCD
	7085	7095	7105	7115	7125	7135
30 EMCR	LYNYGAGIKL	PSGIMLNVVK	YTQLCQYLNS	TTMCVPHNMR	VLHYGAGSDK	GVAPGTTVLK
229E	LYNYGAGLKL	PSGIMFNVVK	YTQLCQYFNS	TTLCVPHNMR	VLHLGAGSDY	GVAPGTAVLK
PEDV	LYNYGAGVKL	PDGIMFNVVK	YTQLCQYLNS	TTMCVPHHMR	VLHLGAGSDK	GVAPGTAVLR
TGEV	LYNYGAQVKL	PDGITTNVVK	YTQLCQYLNT	TTLCVPHHMR	VLHLGAAGAS	GVAPGSTVLR
OV43	LWNYGKPVTL	PTGCMNNVAK	YTQLCQYLNT	TTLAVPVNMR	VLHLGAGSEK	GVAPGSAVLR
BoCoV	LWNYGKPVTL	PTGCMNNVAK	YTQLCQYLNT	TTLAVPVNMR	VLHLGAGSEK	GVAPGSAVLR
MHV	LWNYGKPVTL	PTGCMNNVAK	YTQLCQYLNT	TTLAVPANMR	VLHLGAGSDK	DVAPGSAVLR
40 AIBV	IPNYGVGITL	PSGILMNVAK	YTQLCQYLNS	TTICVPHNMR	VMHFGAGSDK	GVAPGSTVLK
5 SARS CoV	LQNYGENAVI	PKGIMMNVAK	YTQLCQYLNT	LTAVPVNMR	VIHFGAGSDK	GVAPGTAVLR
	7145	7155	7165	7175	7185	7195
45 EMCR	RWLPPD----	-----AIII	DNDINDYVSD	ADFSITGDCA	TVYLEDKFDL	LISDMYDG--
229E	RWLPHD----	-----AIVV	DNDVDYVSD	ADFSVTGDCA	TVYLEDKFDL	LISDMYDG--
PEDV	RWLPLD----	-----AIVV	DNDSDYVSD	ADYSVTGDCA	TVYLEDKFDL	LISDMYDG--
TGEV	RWLPPD----	-----AIVV	DNDLDYVSD	ADFSVTGDCT	SLYIEDKFDL	LVSDLYDG--
50 OV43	QWLPPG----	-----TILV	DNDLYPFVSD	SVATYEGDCI	TLPPDCQWDL	IISDMYDP--
BoCoV	QWLPPG----	-----TILV	DNDLYPFVSD	SVATYEGDCI	TLPPDCQWDL	IISDMYDP--
MHV	QWLPPG----	-----TILV	DNDINPFVSD	SVASYGNCI	TLPIACQWDL	IISDMYDP--
AIBV	QWLPEG----	-----TLLV	DNDIVDYVSD	AHVSVLSDCN	KYNTHEKFDL	VISDMYTDND
5 SARS CoV	QWLPTG----	-----TLLV	DNDLNDYVSD	ADSTLIGDCA	TVHTANKWDL	IISDMYDP--
	7205	7215	7225	7235	7245	7255
55 EMCR	--RIKFCDE	NVSKDGFFTY	LNGVIREKLA	IGGSVAIKIT	EYSWNKLYE	LIQRFATWTL
229E	--RTKAIDGE	NVSKDGFFTY	INGFICEKLA	IGGSIAIKVT	EYSWNKLYE	LVQRFATWTL
PEDV	--KIKSCDGE	NVSKDGFFTY	INGVITEKLA	LGGTVAIKVT	EFSWNKKLYE	LIQKFEYWTM
TGEV	--STKSIDGE	NVSKDGFFTY	INGFIKEKLS	LGGVAIKIT	EFSWNKLYE	LIQKFEYWTM
60 OV43	--ITKNIGY	NVSKDGFFTY	ICHMIRDKLA	LGGVAIKIT	EFSWNAELYK	LMGYFAFWTV
BoCoV	--LLLDIGVH	VVRCS---YI	HCHMIRDKLA	LGGVAIKIT	EFSWNAELYK	LMGYFAFWTV
MHV	--LTKNIGY	NVSKDGFFTY	LCHLIRDKLA	LGGVAIKIT	EFSWNAELYK	LMGYFAFWTV
AIBV	SKRKHEGVIA	NNGNDDVFIY	LSSELRNNLA	LGGFAVKVT	ETSWHEVLYD	IAQDCAWWTM
65 SARS CoV	--RTKHVTKE	NDSKEGFFTY	LCGFIKQKLA	LGGSIKVKIT	EHSWNADLYK	LMGHFSWWTM
	7265	7275	7285	7295	7305	7315
70 EMCR	FCTSVNTSSS	EAFLLIGINYL	GDFIQGPFIA	GNTVHANYIF	WRNSTIMSLS	YNSVLDLSKF
229E	FCTSVNTSSS	EAFVVGGINYL	GDFIQGPFID	GNTVHANYIF	WRNSTIMSLS	YNSVLDLSKF
PEDV	FCTSVNTSSS	EAFLLIGINYL	GDFASGAVID	GNTMHANYIF	WRNSTIMTMS	YNSVLDLSKF
TGEV	FCTSVNTSSS	EGFLIGINYL	GPYCDKAIVD	GNTMHANYIF	WRNSTIMTMS	YNSVLDLSKF
OV43	FCTNANASSS	EGFLIGINYL	CKPKV--EID	GNVMHANYIF	WRNSTVWNGG	AYSFLDMAKF
BoCoV	FCTNANASSS	EGFLIGINYL	GKPKV--EID	GNVMHANYIF	WRNSTVWNGG	AYSFLDMAKF
MHV	FCTNANASSS	EGFLIGINYL	NRTRT--EID	GKTMHANYIF	WRNSTVWNGG	AYSFLDMAKF
75 AIBV	FCTAVNASSS	EAFLLIGINYL	GASEK-VKVS	GKTLHANYIF	WRNCNYLQTS	AYSIFDVAKF
5 SARS CoV	FVTNVNASSS	EAFLLIGINYL	GKPKV--QID	GKTLHANYIF	WRNTNPIQLS	AYSIFDVAKF
	7325	7335	7345	7355	7365	

77/87

EMCR	ECKHKATVVV	TLKDSVDVND	VLSLIKSGRL	LLRNSGRFGG	FSNHLVSTK-
229E	NCKHKATVVV	QLKDSIDINEM	VLSLVRSGKL	LVRNGKCLSL	FSNHLVSTK-
PEDV	NCKHKATVVV	NLKDSSISDV	VLGLLKNKGL	LVRNNDACIG	FSNHLVNVNK
TGEV	KCRCNNALIV	NLKEKELNEM	VIGLLRKGGK	LIRNNGKLLN	FGNHFVNTP-
OV43	PLKLAGTAVI	NLRADQINDM	VYSLLEKGGK	LIRDNTKEVF	VGDSLNVNI-
BoCoV	EIPQFGTGVL	IACLIWLNSR	LSWLVMF---	---	---
MHV	PLKVAGTAVV	SLKPDQINDL	VLSLIEKGGK	LVRDTRKEVF	VGDSLNVNVK-
AIBV	DLRLKATPVV	NLKTEQKTDL	VFNLIKCGKL	LVRDVGNTSF	TSDSFVCTM-
SARS CoV	PLKLRGTAVM	SLKENQINDM	IYSLLEKGGK	IIRENNRVVV	SSDILVNN--

## e. Putative Spike protein

		5	15	25	35	45	55
20	EMCR S	---MKLFLI	LLILP---	L VSCFSTC---	-----N	SNASIS----	-----ML
	229E S	---	---	---	---	---	---
	PEDV	MRSLIYFWLL	LPVLP----	LSLPQDV---	-----T	RCQSTTNF--	-----RREFS
	TGEV	---MKKLFVV	LVMF----	IYGDNFP---	-----C	SKLTNRTIGN	QWNLIETELL
	CaCoV	---MIVLTL	LFLFL-YSSV	SCTSNND---	-----C	VQVNVTLQPG	NENIIKDELFL
25	FeCoV	---MIVLTC	LLLLCSYHTV	LSTTNNE---	-----C	IQVNVTLQAG	NENLIRDFLF
	Por Resp C	---MKKLFVV	LVMF----	IYG-----	---	---	---
	OC43	---MFLIL	LISLPTAFV	IGDLKCTSDT	SYINDKDTGP	PPISTDTVDV	TNGLGTYIVL
	BoCoV	---MFLIL	LISLPMFAV	IGDLKCT--T	VSINDVDTGA	PSISTDIVDV	TNGLGTYIVL
	MHV	---MLFVF	LTLPLSSLG	IGDFRCIQ-L	VNTDTSNASE	PSVSTEVVDV	SKGIGTYIVL
30	Rat CoV	---MLFVF	LTLPLSCLG	IGDFRCIN-L	VNTRISNARA	PSVSTEVVDV	SKGLGTYIVL
	PHEV	---MFFIL	LISLPSAFV	IGDLKCT--T	SLINDVDTGV	PSISSEVVDV	TNGLGTYIVL
	AIBV	---	---	---	---	---	---
	SARS	---MFIFLL	FLTLTSG---	-SDLDL---	---	-CTTFDDVQA	PNYTQHTSSM
35		65	75	85	95	105	115
	EMCR S	QLG--VPDMS	STIVTGLLP-	---VHWICAN	QSTSSYPANG	FFYIDVG-KH	RSAFALHSGY
	229E S	---	---	---	---	---	---
	PEDV	KFN--VQAPA	VVVLGGYLP	MNSSSWYCGT	GIETASGVHG	IFLSYIDSGQ	GFEIGTSQEP
40	TGEV	NYSSRLPPNS	DVVLGDYFPT	--VQPFNCI	RNDSNDLYVT	LENLKALYWD	YATENITWN-
	CaCoV	QN---FKEEG	SLVGGYYP-	--TEVWYNC	TTOQTAYKY	FSNIHAFYFD	MEAMENSTGN
	FeCoV	SN---FKEEG	SVVGGYYP-	--TEVWYNC	RTARTAFQY	FNNIHAFYFV	MEAMENSTGN
	Por Resp C	---	---	---	---	---	---
	OC43	DR---VYLNT	TFLNGXYPT	SGSTYRNMAL	KGSVLLSRLW	FKPPFLSDFI	NGIFAKVKNT
45	BoCoV	DR---VYLNT	TLLNGXYPT	SGSTYRNMAL	KGTLLLSRLW	FKPPFLSDFI	NGIFAKVKNT
	MHV	DR---VYLNA	TLLLTGYYPV	DGSMYRNMAL	TGINTISLW	YKPPFLSEFN	DGIFAKVKNL
	Rat CoV	DR---VYLNA	TLLLTGYYPV	DGSMYRNMAL	MGTNTLSLW	FEPPFLSEFN	DGIYAKVKNL
	PHEV	DR---VYLNT	TLLNGXYPT	SGATFRNMAL	KGTLLLSLW	FKPPFLSDFI	DGIFAKVKNS
	AIBV	---	---	---	---	---	---
50	SARS	RG---VYYPD	EIERSDTLYL	TQDLFLPFYS	NVTGFHTINH	TFGNPVIFPK	DGIYFAATEK
55		125	135	145	155	165	175
	EMCR S	YDANQYIYL	TNKH-----	---	LNAPVTLKIC	KFGN-----	---TSDFLS
	229E S	---	---	---	---	---	---
	PEDV	FDPGSGYQLYL	HKATNG----	-----N	TNAIARLRIC	QFPDN-----	---KTLGPTVN
	TGEV	-HRQRLNVVV	NGYPYSITV-	-----TTRN	FNSAEGAIIC	ICKGSPPTTT	TESSLTCNWG
	CaCoV	ARGKPLLVHV	HGNPVSIIVY	ISAYRDDVQF	RPLLKHGILL	ITKN--DTVD	YNSFTINQWR
60	FeCoV	ARGKPLLEHV	HGEPVSVII-	-SAYRDDVQ	RPLLKHGILVC	ITKN--RHIN	YEQFTSNQWN
	Por Resp C	---	---	---	---	---	---
	OC43	KVIKDRVMYS	EFPATITG--	-----STF	VNTSYSVVVQ	PRTINSTQDG	YNKLQGLLEV
	BoCoV	KVIKGVVMYS	EFPATITG--	-----STF	VNTSYSVVVQ	PHTTN----L	DNKLQGLLEI
	MHV	KASLPKDSIS	YFPTIIIG--	-----SNF	VNTSYTVVLE	PYN-----	---GIIMA
	Rat CoV	KASLPIGSAS	YFPTIIIG--	-----SNF	VNTSYTVVLE	PYN-----	---GIIMA
65	PHEV	RFSKDGVIYS	EFPATITG--	-----STF	VNTSYSIVVE	PHTSL----I	NGNLQGLLQI
	AIBV	---	---	---	---	---	---
	SARS	SNVVRGWVFG	STMNKSQS-	-----VII	INNSTNVVIR	ACN----FEL	CDNPFFAVSK
70		185	195	205	215	225	235
	EMCR S	NVSTSHDCIV	NLSFTEQL--	GVPLGTTISG	ETVRLHLYNA	TRTFYVPAAY	KLTKLSVKCY
	229E S	---	---	---	---	---	---
	PEDV	DVTTRGRNCLF	NKAIPAYMRD	GKDIVVGITW	DNDRTVTF-A	--MFVLLVAY	ALLHIAG---
	TGEV	SECR-LNHKF	PICPSNSEAN	CGNMLYGLQW	FADEVVAYLH	DKIYHFLYKN	DWSRVATRCY
75	CaCoV	DICLGDDRKI	PFSVVPDND	-GTKLEGLEW	NDDYVTAYIS	DESHRLNINN	QWSGTVTFGD
	FeCoV	STCTGADRKI	PFSVIPTDN	-GTKIYGLEW	NDDFVTAYIS	GRSYHLNINT	NWFNNVTLLY
	Por Resp C	---	---	---	---	---	---
	OC43	SVCQYNMCEY	PQTICHPNLG	--NHRKELWH	LDTGVVVSCLY	KRNFTYDVNA	DYLYFHEYQ-
	BoCoV	SVCQYTMCEY	PHTICHPKLG	--NKRVELWH	WDTGVVVSCLY	KRNFTYDVNA	DYLYFHEYQ-
80	MHV	SICQYITICQL	PYTDCPKNTG	G-NKILGFWH	TELKSPVCIL	KRNFTFVNNA	EWLYFHEYQ-

78/07

Rat CoV	SICQYITICQL	PHTDCKPNTG	G-NTLIGFWH	TDLRPPVCIL	KRNFTFNUNA	EWLYFHFYQ-
PHEV	SVCQYTMCEY	PHTICHPNLG	--NQRIELWH	YDTDVVSCLY	RRNFTYDVNA	DYLYFHFYQ-
AIBV	-----	-----	-----	---MLVTPLL	LVTLLCALCS	AVLYDSS---
SARS	PMGTQHTMI	FDNAFNCTFE	YISDAFSLDV	SE-KSGNFKH	LREFVFNKND	GFLYVYKG--
5	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	245	255	265	275	285	295
EMCR S	FSESCWF---	---SVVNA-	-T-ITVNVTT	LNGRIVNYTV	CDD---CNGY	TDNIFSVQQD
229E S	-----	-----	---CQT	TNGLNTSYSV	CNG---CVGY	SENVFAVESG
PEDV	NRRSCAM---	---QYVYTP	TY-YMLNVT	AGEDGIYEP	CTAN---CTGY	AANVFATDSN
TGEV	MRATTLEVAG	TLVDLWVFN	VIDVSYRVN	NKNGTTVSN	CTD---QCASY	VANVFTTQPG
CaCoV	SRTSTATWQH	S---AAYVYQG	VSNFTYYKLN	KTAGLKSYEL	CEDYEYCTGY	ATNVFAPTSG
FeCoV	SRSSTATWEY	S---AAYAYQG	VSNFTYYKLN	NTNGLKTYEL	CEDYEHCTGY	ATNVFAPTSG
Por Resp C	-----	-----	---TSUVSN	CTD---QCASY	VANVFTILPG	
OC43	EGG-----	-----	-TFYAYFTDT	GVVTKFLENV	YLG---MALS	HYVYMLPTCN
BoCoV	EGG-----	-----	-TFYAYFTDT	GVVTKFLENV	YLG---TVLS	HYVYMLPTCS
MHV	QGG-----	-----	-TFYAYADV	SSATTFLESM	YIG---DVL	QYFVLPYMCT
Rat CoV	QGG-----	-----	-TFYAYADV	SSATTFLESM	YIG---AVLT	QYFVLPYMCS
PHEV	EGG-----	-----	-TFYAYFTDT	GVVTKFLENV	YLG---TVLS	HYVYMLPTCN
AIBV	S-----	-----	---YVYYQ	AFRPPSGWHL	QGG---AYAV	VNISSEFNNA
SARS	YQP-----	-----	IDVVRDLPSG	FNTLKPIFKL	PLG---INITNF	RAILTAFFSPA
25	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	305	315	325	335	345	355
EMCR S	GRIPNGFPFN	NWFL-LTNGS	TLVDGVSRLY	QPLRLTCLWP	VPGLKSSTGF	VYFNATGSDV
229E S	GYIPSDFAFN	NWFL-LTNTS	SVVDGVVRSF	QPLLLNCLWS	VSGLRRTTGF	VYFNGTG-RG
PEDV	GHIPEGFSFN	NWFL-LSNDS	TLLHGKVVSN	QPLLVNCLLA	IPKIYGLQGF	FSFNHTM-DG
TGEV	GFIPSDFSFN	NWFL-LTNSS	TLVSGKLVT	QPLLVNCLWP	VPSFEEAAT	FCFEGAG-FD
CaCoV	GYIPDGFSEF	NWFL-LTNSS	TFVSGREVTN	QPLLVNCLWP	VPSFGVAAQE	FCFEGAG-FS
FeCoV	GYIPDGFSEF	NWFL-LTNSS	TFVSGREVTN	QPLLVNCLWP	VPSFGVAAQE	FCFEGAG-FS
Por Resp C	GFIPSDFSFN	NWFL-LTNSS	TLVNGKLVTK	QPLLVNCLWP	VPSFEEVAST	FCFEGAG-FD
OC43	SKVKNGETLE	YVVTPLTSRQ	YLLAFNQDGI	IFNAVDCMSD	FMSEIKCKTQ	SIAPPTG-VY
BoCoV	S-----	AMTLE	YVVTPLTSKQ	YLLAFNQDGV	IFNAVDCMSD	FMSEIKCKTL
MHV	LTTTGVSFPQ	YVVTPLVKRQ	YLFNFNQKGI	ITSVDCASS	YTSEIKCKTQ	SMNPNTG-VY
Rat CoV	PTTSGVSSPQ	YVVTPLVKRQ	YLFNFNQKGI	ITSVDCASS	YTSEIKCKTQ	SMNPNTG-VY
PHEV	S-----	ALSL	YVVTPLTRQ	FLAFDQDGV	LYHAVDCASD	FMSEIMCKTS
AIBV	G-----	SSS	GCTVGIHGG	RVVNASSIAM	TAPSSGMAWS	SSQFCTAHCN
SARS	QDIWGTSA	YFVGYLKPTT	FMLKYDENG	ITDAVDCSQN	PLAELKCSVK	SFEIDKG-IY
40	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	365	375	385	395	405	415
EMCR S	NNGYQHNSV	ADVRYNLNL	SANSVDNLKS	GVIVFKTLQY	-DVLFYCSN-	--SS-SGULD
229E S	DCKGSSDVL	SDVIRYNLNF	EEN---LRR	GTILEKTSYG	-VVVFYCTN-	--NT-LVSGD
PEDV	VCNGAAMDRA	PEALRFNLND	TSV---ILAE	GSIVLHTALG	TNLSFVCSN-	--SSDPLHAI
TGEV	QCNGAVLNNT	VDVIRENLNF	TNNVQSGKGA	TVFSLNTTGG	VTLEISCY--	TVSDSSFFSY
CaCoV	QCNGVSLNNT	VDVIRENLNF	TADVQSGMGA	TVFSLNTTGG	VILEISCYND	TVSESSFFSY
FeCoV	QCNGVSLNNT	VDVIRENLNF	TADVQSGMGA	TVFSLNTTGG	VILEISCYND	TVSESSFFSY
Por Resp C	QCNGAVLNNT	VDVIRENLNF	TNNVQSGKGA	TVFSLNTTGG	VTLEISCYND	TVSDSSFFSY
OC43	ELNGYTVQPI	ADVYRRKLN	PNCNIEAWLN	DKSVSPPLNW	ERKTFSNCF	NMSSLMSFIQ
BoCoV	ELNGYTVQPI	ADVYRRIPNL	PDCNIEAWLN	DKSVSPPLNW	ERKTFSNCF	NMSSLMSFIQ
MHV	DLSGYTVQPV	GLVYRRVRNL	PDCKIEEWLT	AKSVSPPLNW	ERKTFQNCNF	DLSSLRFRVQ
Rat CoV	ELNGYTVQPV	GLVYRRVRNL	PDCKIEEWLT	AKSVSPPLNW	ERKTFQNCNF	DLSSLRFRVQ
PHEV	ELNGYTVQPV	ATVYRRIPDL	PNCNIEAWLN	SKTVSPPLNW	ERKTFSNCF	NMSSLMSFIQ
AIBV	CYKHGGCPLT	GMLQONLIRV	SAMKNGQLFY	NLTVSVAKYP	TFRSFQCVN-	--NLTSVYLN
SARS	QTSNFRVVP	GDVVRFPNIT	NLCPFGEVFN	ATKFPVYAW	ERKKISNCVA	DYSVLNSTF
60	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	425	435	445	455	465	475
EMCR S	TTIPFGPSSQ	PYYCFINSTI	NTHVSTFVG	ILPPTVREIV	VARTGQFYIN	GKXYFDLGF
229E S	AHIFPGTVLG	NFYCFVNTTI	GNETTSFVVG	ALPKTVREFV	ISRTGHEFIN	GYRYFTLGNV
PEDV	FAIPLGATEV	PYYCFVKVD	YNSTVYKFLA	VLPPTVREIV	ITKYGDVYVN	GFGYHLGLL
TGEV	GEIPFGVTDG	PRYCYV---H	YNGTALKYLG	TLPPSVKEIA	ISKWGHFYIN	GYNEFFSTFPI
CaCoV	GEIPFGVTDG	PRYCYV---H	YNGTALKYLG	TLPPSVKEIA	ISKWGHFYIN	GYNEFFSTFPI
FeCoV	GEIPFGVTDG	PRYCYV---L	YNGTALKYLG	TLPPSVKEIA	ISKWGHFYIN	GYNEFFSTFPI
Por Resp C	GEIPFGVTDG	PRYCYV---L	YNGTALKYLG	TLPPSVKEIA	ISKWGHFYIN	GYNEFFSTFPI
OC43	ADSFTCNNID	AAKIYG---MC	FSSITIDKFA	IPNGRKVDLQ	LGNLGYLQSF	NYRIDTTATS
BoCoV	ADSFTCNNID	AAKIYG---MC	FSSITIDKFA	IPNGRKVDLQ	LGNLGYLQSF	NYRIDTTATS
MHV	AESLSCSNID	ASKVYG---MC	FGSISIDKFA	IPNRRRVLDQ	LGNLGYLQSF	NYKIDTRATS
Rat CoV	AESLSCSNID	ASKVYG---MC	FGSISIDKFA	IPNRRRVLDQ	LGNLGYLQSF	NYKIDTRATS
PHEV	ADSFGCNID	ASRLYG---MC	FSSITIDKFA	IPNRRRVLDQ	LGNLGYLQSF	NYKIDTRATS
AIBV	GDLVYTSNET	IDVTSAG---V	YFKAGGPITY	KVMREVKALA	YFVNGTAQDV	ILCDGSPRGL
SARS	FSTFKCYGVS	ATKLN---LC	FSNVYADSFV	VKGDDVRQIA	PGQTGVIADY	NYKLPPDDFMG
75	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	485	495	505	515	525	535
EMCR S	EAVNFNVT--	--TASATDFW	TVAFAFVDV	LVNVSATNIQ	NLLYCDSPFE	KLQCEHLQFG
229E S	EAVNFNVT--	--TAETDFC	TVALASYADV	LVNVSQTSIA	NIIYCNSVIN	RLRCDQLSFD
PEDV	DAVTINFTGH	GTDDDVSGFW	TIASNTFVDA	LIEVQGTISI	RILYCDSPFE	QLKCSQVAFD
TGEV	DCISENLT--	--TGDSDFW	TIATSYTEA	LVQVENTAIT	KVTYCNSHVN	NIKCSQITAN
CaCoV	DCIAFNLT--	--TGASGAFW	TIATSYTEA	LVQVENTAIK	KVTYCNSHIN	NIKCSQITAN
80	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....



79/87

	FeCoV Por Resp C OC43 BoCoV MHV Rat CoV PHEV AIBV SARS	GCISFNLT-- --TGVSAGFW DCISFNLT-- --TGDSDFVW COLYYNLP-- --AANVSVS COLYYNLP-- --AANVSVS COLYYSLA-- --KNNVTVN COLYYSLA-- --QDNVTVI COLYYSLP-- --AANVSVT LACQYNTG-- --NFSDGEY CVLAWNTR-- --NID	TIAYTSYTEA TIAYTSYTEA RFNPSTWNKR RFNPSTWNRR NHNPSWNR NHNPSWNR HYNPSWNR PFTNSLVKQ ATSTGNYNK	LVQVENTA LVQVENTAIT FGFIEDSVFK FGFTEQFVK YGFND---- YGFND---- YGFNN---- KFIVYR-- YRYLR----	NVTYCNSHIN NVTYCNSYVN PRPAGVLTH PQPVGVFTH -VATFGTGKH -VATFHSGEH -QSFGSRGLH -ENSUNT -HG----	NIKCSQLTAN NIKCSQLTAN DVVYAQHC DVVYAQHC DVAYAEACFT DVAYAEACFT DAVYSQCFN TCTLHNFIFH -----	
10		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....	545 555 565 575 585 595				
15	EMCR S 229E S PEDV TGEV CaCoV FeCoV Por Resp C OC43 BoCoV MHV Rat CoV PHEV AIBV SARS	LQDGFY--SA VPDGFY--ST LDDGFYPISS LNNGFYFVSS LQNGFYFVAS LNNGFYFVAS LNNGFYFVSS LNNGFYFVSS APKNFCPCCKL VGASYCPCAN VGASYCPCAK TPNTYCPCRT NETGANPNPS ---KLRFER	--NFLDDNVL --SPIQSVEL --RNLLSHEQ --SEVG--LV --SEVG--LV --SEVG--LV --SEVG--LV --SEVG--LV NGS--CVGSGP DGSCLVGNP P--SIVSPCTT P--STVYSCVT --SQCIG-- -----	P-----ET P-----VS P-----IS N-----KS N-----KS N-----KS N-----KS N-----KS G-----KNNG GIDAGYKNSG G-----K-PN G-----K-PK G-----AG -----D	YVALPIYYQH IVSLPVYHKH FVTLPSFNDH VLLPSFYTH VLLPSFYSH VLLPSFFTY VLLPSFLTH VLLPSFLTH IGTCPAGTNY IGTCPAGTNY FANCPGTSTN SANCPGTSTN TGTCVPVGT VQNIQTYQTK ISNVPFSPDG	TDINFATA-- TFIVLYVDFK SFVNITVSA-- TIVNITIGLG TSVNITIDLG TAVNITIDLG TIVNITIGLG LTCND---- LTCNNA-- RECTVMPLAN RECNAVQASG RKCFAAVTK-- TAQSGYNNFN KPCTP----	--SFGGSCY PQSGGGKCFN --AFGG--LSS --MKRSGYGQP --MKRS--VTVT --MKLSGYGQP --MKRSGYGQP -----LC -----LC --NQFKDCTC --FKSKDCTC --ATKCTCWC -----FSF
30	EMCR S 229E S PEDV TGEV CaCoV FeCoV Por Resp C OC43 BoCoV MHV Rat CoV PHEV AIBV SARS	CKPRQVNISL CYPAGVNITL ANLVASDTTI IASTLSNITL IASPLSNITL IASTLSNITL IASTLSNITL IASTLSNITL TPDPITFKAT TPDPITSKST NPSPLTTYDL NPSPLTTYDP QPDPSYKGV LSSFVYKESN -----	N---GNTSV ANFNETRGFL N---GFSSF PMQDHNITDVY PMQDNNITDVY PMQDNNITDVY PMQDNNITDVY PMQDNNITDVY GTYPKQTKS GPKYKQTKS --R--CLQARS --R--CLQARS NAWTCPOQSKV FMYGSYHPSC -----P--AL	CVRTSHFSIR CVDTSHFTTK CVDTQRQFTT CIRSDQFSVY CIRSNQFSVY CIRSNQFSVY CIRSNQFSVY CIRSDQFSVY LVGIGEHCSG LVGIGEHCSG MLGVGDHCEG MLGVGDHCEG SIQPGQHCPCG KFRLETINNG NCYWELNDY-	YIYNRVKSGS YVAVYANVG-- LFYNVTNSYG VHSTCKSALW VHSTCKSSLW VHSTCKSSLW VHSTCKSSLW VHSTCKSVLW LAVKSDYCGG LAIKSDYCGG LGLEDKCGG LGLEDKCGG LGLVEDDCSG LWENSLSVS- -----G	P-----G ----- ----- DNIFKRNCTD DNNFNACTD DNIFNQDCTD DNIFKRNCTD -----N -----N S-----N S-----N -----N ----- -----	655 DSSWHIYLKS --RWSASINT -----YVSKSQD VLDATAVIKT VLDATAVIKT VLDATAVIKT VLDATAVIKT SCTCRPQAF PCTCQPAFL TCNCSAHAFV ICNCSADAFV PCTCKPQAFI --IAYGPLQ --FYTTTGI
45	EMCR S 229E S PEDV TGEV CaCoV FeCoV Por Resp C OC43 BoCoV MHV Rat CoV PHEV AIBV SARS	GTCPFFSFSKL GNCPEFSEFQV SNCPFTLQSV GTCPFFSFDKL GTCPFFSFDKL GTCPFFSFDKL GTCPFFSFDKL GWSADSCLOQ GWSVDSCLQG GWAKDSCLAN GWAMDSCLSN GWSSETCLQN GGCKQSVFKG GYQPYRVVVL	NNFQKFKTIC NNFVKFGSVC NDYLSFSKFC NNYLTFNKFC NNYLTFNKFC NNYLTFNKFC NNYLTFNKFC DKCNIFANFI DRCNIFANFI GRCHIFSNLM ARCHIFSNLM GRCNIFANFI RATCCYAYS S-----FELLN	FSTVEVPGSC FSLKDIPEGG VSTSLLAGAC LSLSPVGANC LSLNPVGANC LSLSPVGANC LSLSPVGANC LHDVNSGLTC FHDVNSGTTT LNGINSGLTC LNGINSGLTC LNDVNSGTTT GGPSLCKGVY APATVCGPKL	NFPLEATW-- AMPIVANW-- TIDLFGYP-- KFDVAAR-- KFDVAAR-- KFDVAAR-- KFDVAAR-- STDLOKANTD STDLOKSNITD SMDLQLENTE STDLOKANTD STDLOKANTD SGELDHN-- STDLLKN--	HYTSYTIIVA AYSKYYTIGS AFSGSVKLTS TRTNEQVVR TRTNEQVVR TRTNEQVVR TRTNEQVVR IILGVCVNYD IILGVCVNYD VVTGVCVKYD VVTGVCVKYD ITTDVCVNYD ---FECGLLV ---QCVNFN	LYVTWSEGN LYVSWSDGDG LYFQFTKGL LYVIYEEGDN LYVIYEEGDN LYVIYEEGDN LYVIYEEGDN LYGILGQGI LYGILGQGI LYGITGQGI LYGITGQGI LYGITGQGI YVTKSGGSRI FNGLTGTG-V
65	EMCR S 229E S PEDV TGEV CaCoV FeCoV Por Resp C OC43 BoCoV MHV Rat CoV PHEV AIBV SARS	ITGVFPVPSG ITGVFPQVEG ITGTPKPLEG IVGVPSDMSG IVGVPSDMSG IVGVPSDMSG IVGVPSDMSG VEVNATYYNS VEVNATYYNS KEVKADYYNS KEVKADYYNS IEVNATYYNS QTATEPPVIT LTPSSKRFPQ	IREFSNLVLN VSSEFMVTLN ITDVSFMTLD VHDLSVLHLD LHDLVLHLD LHDLVLHLD LHDLVLHLD WQNLLYDSNG WQNLLYDSNG WQNLLYDVNG WQNLLYDVNG WQNLLYDSSG QNNYNNITLN EQQFGRDVS	NCTKYNIYDY KCTKYNIYDY VCTKYTIYGF SCTDYNIGR SCTDYNIGR SCTDYNIGR SCTDYNIGR NLYGFRDIY NLYGFRDIY NLYGFRDIY NLYGFRDIY NLYGFRDIY NLYGFRDIY NLYGFRDIY	VGTGIIRSSN SGVGIVRVSN KGEGLITLTN TGVGIIIRQT TGVGIIIRKT TGVGIIIRRT TGVGIIIRQT NRTFMIRSCY NRTFMIRSCY NRTFMIRSCY NRTFMIRSCY NRTFMIRSCY NRTFMIRSCY NRTFMIRSCY	QSLAGGITYV DTFLNGITYT SSILAGVYIT RTLLSGLYYT STLLSGLYYT STLLSGLYYT STLLSGLYYT SGRVSAAFHA SGRVSAAFHA SGRVSAAYHQ SGRVSAAYHQ SGRVSAVFAH DSAVSYNYLA FEGVSVITPG	S----- S----- S----- S----- S----- S----- S----- N----- N----- D----- D----- N----- N----- A-----
75	EMCR S 229E S PEDV TGEV CaCoV FeCoV Por Resp C OC43 BoCoV MHV Rat CoV PHEV AIBV SARS	ITGVFPVPSG ITGVFPQVEG ITGTPKPLEG IVGVPSDMSG IVGVPSDMSG IVGVPSDMSG IVGVPSDMSG VEVNATYYNS VEVNATYYNS KEVKADYYNS KEVKADYYNS IEVNATYYNS QTATEPPVIT LTPSSKRFPQ	IREFSNLVLN VSSEFMVTLN ITDVSFMTLD VHDLSVLHLD LHDLVLHLD LHDLVLHLD LHDLVLHLD WQNLLYDSNG WQNLLYDSNG WQNLLYDVNG WQNLLYDVNG WQNLLYDSSG QNNYNNITLN EQQFGRDVS	NCTKYNIYDY KCTKYNIYDY VCTKYTIYGF SCTDYNIGR SCTDYNIGR SCTDYNIGR SCTDYNIGR NLYGFRDIY NLYGFRDIY NLYGFRDIY NLYGFRDIY NLYGFRDIY NLYGFRDIY NLYGFRDIY	VGTGIIRSSN SGVGIVRVSN KGEGLITLTN TGVGIIIRQT TGVGIIIRKT TGVGIIIRRT TGVGIIIRQT NRTFMIRSCY NRTFMIRSCY NRTFMIRSCY NRTFMIRSCY NRTFMIRSCY NRTFMIRSCY NRTFMIRSCY	QSLAGGITYV DTFLNGITYT SSILAGVYIT RTLLSGLYYT STLLSGLYYT STLLSGLYYT STLLSGLYYT SGRVSAAFHA SGRVSAAFHA SGRVSAAYHQ SGRVSAAYHQ SGRVSAVFAH DSAVSYNYLA FEGVSVITPG	S----- S----- S----- S----- S----- S----- S----- N----- N----- D----- D----- N----- N----- A-----
80		.... ....  .... ....  .... ....  .... ....  .... ....  .... ....	785 795 805 815 825 835				

5	EMCR S 229E S PEDV TGEV CaCoV FeCoV Por Resp C OC43 BoCoV MHV Rat CoV PHEV AIBV SARS	NSGNLLGFKN VSTGNIFIVT PCNQPDQVAV YQQ-SIIGAM TAVNESRYGL QNLLQLPNFY TSGNLLGFKD VTKGTIYSIT PCNPPDQLVV YQQ-AVVGAM LSENFTSYGF SNVVLPKFFF DSGQLLAFKN VTSGAVYSVT PCSFSEQAAY VND-DIVGVI SLSNS--TF NNTRELPGFF LSGDLGFKN VSDGVIYSVT PCDVSAQAAY IDG-AIVGAM TSINSELLGL THWTTTPNFY LSGDLGFKN VSDGVIYSVT PCDVSAQAAY IDG-AIVGAM TSINSELLGL THWTTTPNFY LSGDLGFKN VSDGVIYSVT PCDVSAQAAY IDG-AIVGAM TSINSELLGL THWTTTPNFY LSGDLGFTN VSDGVIYSVT PCDVSAQAAY IDG-TIVGAI TSINSELLGL THWTTTPNFY SSEPALLFRN IKCNYVFNN S LTRQLQPINY FDS-YLGCVV NAYNSTAISV QTCDLTVGSG SSEPALLFRN IKCNYVFNN S LSRQLQPINY FDS-YLGCVV NADNSTSSVV QTCDLTVGSG APEPALLYRN LKCDYVFNNN ISREETPLNY FDS-YLGCVV NADNSTEAV DACDLRMGSG APEPALLYRN LKCDYVFNNN ISREETPLNY FDS-YLGCVI NADNSTEQSV DACDLRMGSG SSEPALLFRN LKSHVFNNT ILRQIQLVNY FDS-YLGCVV NAYNNTASAV STCDLTVGSG GSIDIFVVGQ EYGLNYYKVN PCEDVNNQFV VSGGKLVGIL TSRNETGSQL LENQFYIKIT SSEVAVLYQD VNCTDVSTAI HADQLTPAWR IYS-TGNNVF QTOAGCLIGA EHVDTSYECD
10		.... ....  .... ....  .... ....  .... ....  .... ....  845 855 865 875 885 895
20	EMCR S 229E S PEDV TGEV CaCoV FeCoV Por Resp C OC43 BoCoV MHV Rat CoV PHEV AIBV SARS	YVS-----NG GNN----- CTTAV MIYSNFGICA DGSILPVRPR NSSDNGISAI YAS-----NG TYN----- CTDV LTVSSFGVCA DGSIAVQPR NVSYDSVSAI YHS-----ND GSN----- CTEPV LVYSNIGVCK SGSIGYV-PS QYGQVKIAPT YYSI---YNY TNDRTGTAI DSNDVDCEPV ITYSNIGVCK NGAFVFIN-V THSDGDVQPI YYSI---YNY TNVMNRGTAI D-NDIDCEPI ITYSNIGVCK NGALVFIN-V THSDGDVQPI YYSI---YNY TSERTGTAI DSNDVDCEPV ITYSNIGVCK NGALVFIN-V THSDGDVQPI YYSI---YNY TNDKTRGTPI GSNDVDCEPV ITYSNIGVCK NGALVFIN-V THSDGDVQPI YCVD---YSK NRR----- SRGAI TTGYRETNFE PFTVNSVN-- --DSLEPVG YCVD---YST KRR----- SRAI TTGYRETNFE PFTVNSVN-- --DSLEPVG LCVN---YST SHR----- ARSSV STGYKLTFE PFTVRIVN-- --DSVESVD LCVN---YSI AHR----- ARRSV STGYKLTFE PFTVSIVN-- --DSVESVD YCVD---YVT ALR----- SRRSF TTGYRETNFE PFAANLVN-- --DSIEPVG NGTRRFRRSI TEN----- VANCPY VSYGKFCIKP DGSIAIVPK QLEQFVAPLF IPIG---AGI CAS----- YHTVSL LRSTSQKSIV AYTMSLG--- --ADSSIAY
35	EMCR S 229E S PEDV TGEV CaCoV FeCoV Por Resp C OC43 BoCoV MHV Rat CoV PHEV AIBV SARS	.... ....  .... ....  .... ....  .... ....  .... ....  905 915 925 935 945 955 -ITANLSIPS NWTTSVQVEY LQITSTPIVV DCATYVCNNG PRCKNLLKQY TSACKTIEDA -VTANLSIPS NWTTSVQVEY LQITSTPIVV DCSTYVCNNG VRCVELLKQY TSACKTIEDA -VTGNISIPT NFSMSIRTEY LQLYNTFVSV DCATYVCNNG SRCKQLLTQY TAACKTIESA -STGNVTIPT NFTISVQVEY IQVYTPVSI DCSRYVCNNG PRCKNLLTQY VSACQTIEQA -STGNVTIPT NFTISVQVEY IQVYTPVSI DCARYVCNNG PRCKNLLTQY VSACQTIEQA -STGNVTIPT NFTISVQVEY IQVYTPVSI DCARYVCNNG PRCKNLLTQY VSACQTIEQA -STGNVTIPT NFTISVQVEY IQVYTPVSI DCSRYVCNNG PRCKNLLTQY VSACQTIEQA -GLYEIQIPS EFTIGNMEEF IQTSSPKVTI DCAAFVCGDY AACKSQLVEY GSFCNDINAI -GLYEIQIPS EFTIGNMEEF IQTSSPKVTI DCSAFVCGDY AACKSQLVEY GSFCNDINAI -GLYELQIPT NFTIASHQEF VQTRSPKVTI DCAAFVCGGH TACRQQLVYD GSFCNDINAI -GLYELQIPT NFTIASHQEF VQTRSPKVTI DCAAFVCGDY TACRQQLVYD GSFCNDINAI -GLYEMQIPT NFTIASHQEF IQTRSPKVTI DCAAFVCGDY TACRQQLAEX GSFCNDINAI -GLYEIQIPS EFTIGNLEEF IQTRSPKVTI DCAAFVCGDY AACRQQLAEX GSFCNDINAI NVTENVLIPI SFNLVTDEY IQTRMDKVQI NCLQYVCGSS LDCRKLFOQY GPVCDNLSV -SNNTIAIPT NFSISITTEV MPVSMAKTSV DCNMYICGDS TECANLLQY GSFCQNLNRA
50	EMCR S 229E S PEDV TGEV CaCoV FeCoV Por Resp C OC43 BoCoV MHV Rat CoV PHEV AIBV SARS	.... ....  .... ....  .... ....  .... ....  .... ....  965 975 985 995 1005 1015 LRLSAHLETN DVSSMLTFDS NA-FSLANVT SFG-----D YNLSSVLPQ- LRNSARLESA DVSEMLTFDK KA-FTLANVS SFG-----D YNLSSVIPS- LQLSARLESV EVNSMLTISE EA-LQLATIS SFNG-----DG YNETNVLGAS- LAMGARLENM EVDSMLFVSE NA-LKLASVE AFN-----SS ETLDPIYKEW PNIGGSWLEG LAMGARLENM EIDSMFLVSE NA-LKLASVE AFN-----ST ENLDPIYKEW PNIGGSWLEG LAMGARLENM EVDSMLFVSE NA-LKLASVE AFN-----ST ENLDPIYKEW PNIGGSWLEG LAMGARLENM EVDSMLFVSE NA-LKLASVE AFN-----SS ETLDPIYKEW PNIGGSWLEG LTVNELLDLT TQLQVANSIM NG-VTLSTKL KDGVNFNVD DD INFSPVLGCL G- LTVNELLDLT TQLQVANSIM NG-VTLSTKL KDGVNFNVD DD INFSPVLGCL G- LGEVNNLIDT MQLQVASALI QG-VTLSSRL SDGIGGQIDD INFSPVLGCL G- LGEVNNLIDT MQLQVASALI QG-VTLSSRL ADGIGGQIDD INFSPVLGCL G- LTVNELLDLT TQLQVANSIM NG-VTLSTKI KDGINFNVDD INFSPVLGCL G- VNSVGQKEDM ELLNFYSSTK PAGFNTPVLS NVSTG-----E FNISLLLTN- LSGIAAEQDR NTREVFAQVK QM-YKTPTLK YFG-----G FNFSQILPDP
65		.... ....  .... ....  .... ....  .... ....  .... ....  1025 1035 1045 1055 1065 1075
70	EMCR S 229E S PEDV TGEV CaCoV FeCoV Por Resp C OC43 BoCoV MHV Rat CoV PHEV	----RNHSS RIAGRSAIED LLFSKVVTSG LGTVDDVDYKS CTKGLS--IA DLACAQYNG ----LPTSGS RVAGRSAIED ILFSKLVTS LGTVDDADYK CTKGLS--IA DLACAQYNG V--YDPASGR VVQKRSVIED LLFNKVVTNG LGTVDEDEYK CSNGRS--VA DLVCAQYNG LKYILPSHNS KRKYRSAIED LLFDKVVTS LGTVDEDEYK CTGGYD--IA DLVCAQYNG LKDILPSHNS KRKYRSAIED LLFDKVVTS LGTVDEDEYK SAGGYD--IA DLVCAQYNG LKDILPSHNS KRKYRSAIED LLFDKVVTS LGTVDEDEYK CTGGYD--IA DLVCAQYNG LKYILPSDNS KRKYRSAIED LLFSKVVTSG LGTVDEDEYK CTGGYD--IA DLVCAQYNG -SECSKASS- ----RSAIED LLFDKVKLSD VG-FVEAYNN CTGGAE--IR DLICVQSYNG -SACNKVSS- ----RSAIED LLFSKVVLSD VG-FVEAYNN CTGGAE--IR DLICVQSYNG -SDCGEVTMA AQTGRSAIED VLFDKVKLSD VG-FVEAYNN CTGGQ--VR DLICVQSYNG -SDCSEGTKA AQ-GRSAIED VLFDKVKLSD VG-FVESYNN CTGGQ--VR DLICVQSYNG -SECNRAST- ----RSAIED LLFDKVKLSD VG-FVQAYNN CTGGAE--IR DLICVQSYNG
80		



	AIBV SARS	-----PSSRR K---RSLIED LLFTSVESVG LP-TNDAYKN CTAGPLGFFK DLACAREYNG -----LKPTK- -----RSFTED LLENKVTEAD AG-FMKQYGE CLGDIN--AR DLICAQKFN
5	EMCR S 229E S PEDV TGEV CaCoV FeCoV Por Resp C OC43 BoCoV MHV Rat CoV PHEV AIBV SARS	..... ..... ..... ..... ..... ..... ..... .....  1085 1095 1105 1115 1125 1135 IMVLPGVADA ERMAMVTGSL IGGMVLGGIT S----AAAI FSLALQARLN YVALQTDVLQ IMVLPGVADA ERMAMVTGSL IGGIALGGIT S----AVSI FSLAIQARLN YVALQTDVLQ VMVLPGVVDA EKLHMYASL IGGMALGGIT A----AAAL FSYAVQARLN YLALQTDVLQ IMVLPGVANA DKMTMYTASL AGGITLGALG GG---AVAI FAVAVQARLN YVALQTDVLN IMVLPGVANA DKMTMYTASL TGGITLGALS GG---AVAI FAVAVQARLN YVALQTDVLN IMVLPGVANA DKMTMYTASL AGGITLGALG GG---AVAI FAVAVQARLN YVALQTDVLN IMVLPGVANA DKMTMYTASL AGGITLGALG GG---AVAI FAVAVQARLN YVALQTDVLN IKVLPPLLSE NQISGYTLAA TSASLFPPLT A----AAGV FYLNQYQYRIN GLGVTMDVLS IKVLPPLLSE NQISGYTLAA TSASLFPPLT A----AAGV FYLNQYQYRIN GIGVTMDVLS IKVLPPLLSE NQISGYTAGA TVSAMFP-WS A----AAGV FSLSVQYRIN GLGVTMNVLS IKVLPPLLSE NQISGYTLAA TAASLFPFPT A----AAGV FSLSVQYRIN GLGVTMNVLS LLVLPPIITA EMQALYTSSL VASMAFGGIT A----AGAIP FYLNQYQYRIN GLGVTMDVLS LTVLPPLLTD DMIAAYTAAL VSGTATAGWT FGAGAAQIP FAMQMAYREN HLGITQSLLL GIGVTQNVLY
20	EMCR S 229E S PEDV TGEV CaCoV FeCoV Por Resp C OC43 BoCoV MHV Rat CoV PHEV AIBV SARS	..... ..... ..... ..... ..... ..... ..... .....  1145 1155 1165 1175 1185 1195 ENQKILAAAF NKAINNIVAS FSSVNDIAITH TAEAIHTVTI ALNKIQDVVN QQGSALNHLT ENQKILAAAF NKAMTNIWDA FTGVNDAITQ TSQALQTVAT ALNKIQDVVN QQGNSLNHLT RNQQLAAEF NSAIGNITSA FESVKEAISQ TSKGLNTVAH ALTQVQEVVN QQGSALNHLT KNQQLAASAF NQAIGNITQS FGKVNDAIHQ TSRGLATVAK ALAKVQDVVN IQGQALSHLT KNQQLANAF NQAIGNITQA FGKVNDAIHQ TSKGLATVAK ALAKVQDVVN TQGQALSHLT KNQQLANAF NQAIGNITQA FGKVNDAIHQ TSQGLATVAK ALAKVQDVVN TQGQALSHLT KNQQLAASAF NQAIGNITQS FGKVNDAIHQ TSRGLTTVAK ALAKVQDVVN TQGQALSHLT QNQKLIANAF NNALDAIQEG FDATN-----S ALVKIQAVVN ANAEALNNLL QNQKLIANAF NNALDAIQEG FDATN-----S ALVKIQAVVN ANAEALNNLL ENQKMIASAF NNAIGAIQEG FAATN-----S ALAKMQFVVN ANAEALNNLL ENQKMIASAF NNAIGAIQEG FDATN-----S ALAKIQSVVN ANAEALNNLL QNQKLIASAF NNALDAIQEG FDATN-----S ALVKIQAVVN ANAEALNNLL KNQEKIAASF NKAIGHMQEG FRSTS-----L ALQIQDVVS QKSAILTETM ENQKQIANQF NKAISQIQES LTTTS-----T ALGKLQDVVN QNAQALNTLV
40	EMCR S 229E S PEDV TGEV CaCoV FeCoV Por Resp C OC43 BoCoV MHV Rat CoV PHEV AIBV SARS	..... ..... ..... ..... ..... ..... ..... .....  1205 1215 1225 1235 1245 1255 SQLRHNFQAI SNSIHAIYDR LDSIQADQOV DRLITGRLLA LNAFVSQVLN KYTEVRGSRR SQLRHNFQAI SSSIQAIYDR LDTIQADQOV DRLITGRLLA LNMFVSHLT KYTEVRASRQ VQLQNNFQAI SSSIDDIYSR LDILSADQOV DRLITGRLLA DRLITGRLLA KYTEVQASRK VQLQNNFQAI SSSIDDIYNR LDELSADAQV DRLITGRLLA LNAFVSQTLT ROAEVRASRQ VQLQNNFQAI SSSIDDIYNR LDELSADAQV DRLITGRLLA LNAFVSQTLT ROAEVRASRQ VQLQNNFQAI SSSIDDIYNR LDELSADAQV DRLITGRLLA LNAFVSQTLT ROAEVRASRQ VQLQNNFQAI SSSIDDIYNR LDELSADAQV DRLITGRLLA LNAFVSQTLT ROAEVRASRQ QQLSNRFGAI SSSIQEILSR LDALAEAAQI DRLINGRLTA LNAYVSQQLS DSTLVKFSAA QQLSNRFGAI SSSIQEILSR LDALAEAAQI DRLINGRLTA LNAYVSQQLS DSTLVKFSAA QQLSNRFGAI SASLQEILSR LDALAEAAQI DRLINGRLTA LNAYVSQQLS DMTLVKVSAA QQLSNRFGAI SASLQEILSR LDALAEAAQI DRLINGRLTA LNAYVSQQLS DMTLVKVSAA ASLNKNFGAI SSVIQEIYQQ FDATQANAQV DRLITGRLLS LSVLASAKQA EYIRVSQORE QQLSSNFGAI SSVIQEILSR LDKVEAEVQI DRLITGRLLS LQTVVTQQLI RAAEIRASAN
55	EMCR S 229E S PEDV TGEV CaCoV FeCoV Por Resp C OC43 BoCoV MHV Rat CoV PHEV AIBV SARS	..... ..... ..... ..... ..... ..... ..... .....  1265 1275 1285 1295 1305 1315 LAQQKINECV KSQSNRYGFC G-NGTHIFSI VNSAPDGLLE LHTVLLPTDY KNVKAWSGIC LAQQKINECV KSQSKRYGFC G-NGTHIFSI VNAAPDGLLE LHTVLLPTDY KDVEAWSGLC LAQQKINECV KSQSQRFGFC G-NGTHIFSL VQAAPQGLLE LHTVLVPGDF VNVLAIAGLC LAKDKVNECV RSQSQRFGFC G-NGTHIFSL ANAAPNGMIF FHTVLLPTAY ETVTAWPGIC LAKDKVNECV RSQSQRFGFC G-NGTHIFSL ANAAPNGMIF FHTVLLPTAY ETVTAWSGIC LAKDKVNECV RSQSQRFGFC G-NGTHIFSL ANAAPNGMIF FHTVLLPTAY ETVTAWSGIC LAKDKVNECV RSQSQRFGFC G-NGTHIFSL ANAAPNGMIF FHTVLLPTAY ETVTAWSGIC QAMEKVNECV KSQSSRINFC G-NGNHISL VQNAFYGLYF IHFSYVPTKY VTARVSPGLC QAMEKVNECV KSQSSRINFC G-NGNHISL VQNAFYGLYF IHFSYVPTKY VTAKVSPGLC QAEIKVNECV KSQSSRINFC G-NGNHISL VQNAFYGLYF IHFSYVPTSF TTANVSPGLC QAEIKVNECV KSQSSRINFC G-NGNHISL VQNAFYGLYF IHFSYVPTSF TTANVSPGLC QAEIKVNECV KSQSSRINFC G-NGNHISL VQNAFYGLYF IHFSYVPTSF TTANVSPGLC LATQKINECV KSQSIRYSFC G-NGRHVLT PQNAPNGIVF IHFSYTPDSF VNVTAVGFC LAATKMSECV LGQSKRVDFC G-KGYHLMSE PQAAPHGVVF LHVTVVPFQE RNETTAPATC
75	EMCR S 229E S PEDV TGEV CaCoV FeCoV Por Resp C	..... ..... ..... ..... ..... ..... ..... .....  1325 1335 1345 1355 1365 1375 VDG-----IYG YVLRQPNLVL YS-----DN GVFRVTSRVM FQPRVPLSD FVQIYNENVT VDG-----TNG YVLRQPNLAL YK-----EG NYRITSRIM FEPRIPTMAD FVQIENENVT VNG-----EIA LTLREPLVL FTHLQTYTA TEYFVSSRRM FEPKPTVSD FVQIESCVVT ASDG-DRTFG LVVKDVQLTL FRN-----LD DKFYLTPTMT YQPRVATSSD FVQIEGCDVL ASDG-SRTFG LVVEDVQLTL FRN-----LD EKFYLTPTMT YQPRVATSSD FVQIEGCDVL ASDG-DRTFG LVVKDVQLTL FRN-----LD DKFYLTPTMT YQPRVATSSD FVQIEGCDVL ALDV-DRTFG LVVKDVQLTL FRN-----LD DKFYLTPTMT YQPRVATSSD FVQIEGCDVL
80	EMCR S 229E S PEDV TGEV CaCoV FeCoV Por Resp C	

82/87

5	OC43	IAG----	DRG	IAPKSGYFVN	VN-----	NTWMTGSGY	YYPEPITENN	VVVMSTCAVN
	BoCoV	IAG----	DRG	IAPKSGYFVN	VN-----	NTWMTGSGY	YYPEPITGNN	VVVMSTCAVN
	MHV	ISG----	DRG	LAPKAGYFVQ	DD-----	GEWKFTGSNY	YYPEPITDKN	SVMSSCAAN
	Rat CoV	ISG----	DRG	LAPKAGYFVQ	DH-----	GEWKFTGSNY	YYPEPITDKN	SVMSSCAVN
	PHEV	IAG----	DIG	ISPKSGYFIN	VN-----	NSWMTGSSY	YYPEPITQNN	VVVMSTCAVN
	AIBV	VKPNASQYA	IVPANGRGIF	IQVN-----	GSYYITARDM	YMPRAITAGD	VVTLTSCQAN	
	SARS	HEG-----	KA	YFPREGVEVF	NG-----	TSWFITQRNF	FSPQIITTDN	TFVSGNCDVV
10		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		1385	1395	1405	1415	1425	1435	
	EMCR S	FVNISRVELH	TVIP-DYVDV	NKTLQEFQON	L-PKYVKPNF	DLTFPNLTYL	NLSSELKQLE	
	229E S	FVNISRSELO	TIVP-EYIDV	NKTLQELSYK	L-PNYTVPDL	VVEQYNQITL	NLTSEISTLE	
	PEDV	YVNLTSDQLP	DVIP-DYIDV	NKTLDEILAS	L-PNRTGPSL	PLDVFNATYL	NLTGEIADLE	
	TGEV	FVNATVSDLP	SIIP-DYIDI	NQTVQDILEN	FRPNWTVPEL	TFDIFNATYL	NLTGEIDDDLE	
15	CaCoV	FVNGTVIELP	SIIP-DYIDI	NQTVQDILEN	FRPNWTVPEL	PLDIFHATYL	NLTGEIDDDLE	
	FeCoV	FVNATVIDLP	SIIP-DYIDI	NQTVQDILEN	YRPNWTVPEF	TLDIFNATYL	NLTGEIDDDLE	
	Por Resp C	FVNTTVSDLP	SIIP-DYIDI	NQTVQDILEN	FRPNWTVPEL	TLDVFNATYL	NLTGEIDDDLE	
	OC43	YTKAPYVMLN	TSIP-NLPDF	KEELDQWFKN	QTSVAPDLSL	DY--INVTF	DLQVEMN---	
	BoCoV	YTKAPDVMLN	ISTP-NLHDF	KEELDQWFKN	QTSVAPDLSL	DY--INVTF	DLQVEMN---	
	MHV	YTKAPEVFLN	TSIP-NLPDF	KEELDKWFKN	QTSIAPDLSL	DFEKLNVTL	DLTDEM---	
20	Rat CoV	YTKAPEVFLN	TSIT-NLPDF	KEELDKWFKN	QTSIVPDLSF	DIGKLVNVT	DLSEYEM---	
	PHEV	YTKAPDLMLN	TSTP-NLPDF	KEELYQWFKN	QSSVAPDLSL	DY--INVTF	DLQVEMN---	
	AIBV	YVSVNKTUIT	TFVDNDDFDF	NDELKQWWD	T--KHELPDF	DKFNYTVPIL	DIDSEID---	
	SARS	IGIINNTVYD	PLQP-ELDSF	KEELDKYFKN	HTSPDVLGD	ISG--INASV	NIQKEID---	
		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		1445	1455	1465	1475	1485	1495	
30	EMCR S	AKTASLFQTT	VELQGLIDQI	NSTYVDLKL	NRFENYIKWP	WWVWLIISV	FVVLLSLV	
	229E S	NKSAELNYTV	QKLQTLIDNI	NSTLVDLKWL	NRVETIYKWP	WWVWLCISV	LIEFVSMILL	
	PEDV	QRSESLRNTT	EELRSLINNI	NNTLVDLLEW	NRVETIYKWP	WWVWLIIVV	LIEFVSLV	
	TGEV	FRSEKLHNTT	VELAILIDNI	NNTLVNLEWL	NRIETIYKWP	WYVWLLIGLV	VIFCIPILLF	
	CaCoV	FRSEKLHNTT	VELAILIDNI	NNTLVNLEWL	NRIETIYKWP	WYVWLLIGLV	VIFCIPILLF	
	FeCoV	FRSEKLHNTT	VELAILIDNI	NNTLVNLEWL	NRIETIYKWP	WYVWLLIGLV	VIFCIPILLF	
35	Por Resp C	FRSEKLHNTT	VELAILIDNI	NNTLVNLEWL	NRIETIYKWP	WYVWLLIGLV	VIFCIPILLF	
	OC43	-----	-RLQEAIKVL	NQSYINLKDI	GTYEYVYKWP	WYVWLLICLA	GVAMLVLLFF	
	BoCoV	-----	-RLQEAIKVL	NQSYINLKDI	GTYEYVYKWP	WYVWLLIGFA	GVAMLVLLFF	
	MHV	-----	-RIQDAIKKL	NESYINLKDV	GTYEMYVYKWP	WYVWLLIGLA	GVAVCVLLFF	
	Rat CoV	-----	-RIQDAIKNL	NESYINLKDI	GTYEMYVYKWP	WYVWLLIGLA	GVAVCVLLFF	
	PHEV	-----	-RLQEAIKVL	NQSYINLKDI	GTYEYVYKWP	WYVWLLIGLA	GVAMLVLLFF	
40	AIBV	-----	-RIQGVIIQGL	NDSLIDLEKL	SILKTYIKWP	WYVWLAIIFA	TIIFILILGW	
	SARS	-----	-RLNEVAKNL	NESLIDLQEL	GKYEYVYKWP	WYVWLGFIA	LIAVVMVTIL	
45		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		1505	1515	1525	1535	1545		
	EMCR S	CCLSTGCCGC	CNCLTSSMRG	CCDCGSTKLP	YYEFKVVHVQ	-----	-----	
	229E S	CCCSTGCCGF	FSCFASSIRG	CCES--TKLP	YYDVEKIHQ	-----	-----	
	PEDV	CCISTGCCGC	CGCCGACFSG	CCRG--PRLQ	YEAPEKVVHVQ	-----	-----	
	TGEV	CCCSTGCCGC	IGCLGSCCHS	ICSR--ROFEN	YEPIEKVVHV	-----	-----	
50	CaCoV	CCCSTGCCGC	IGCLGSCCHS	ICSR--QFES	YEPIEKVVHV	-----	-----	
	FeCoV	CCFSTGCCGC	IGCLGSCCHS	ICSR--RQFEN	YEPIEKVVHV	-----	-----	
	Por Resp C	CCCSTGCCGC	IGCLGSCCHS	IFSR--ROFEN	YEPIEKVVHV	-----	-----	
	OC43	ICCCTG-CG-	-TSCFKKCGG	CCDDYTGQOE	LVIKT---SH	DD-----	DD-----	
	BoCoV	ICCCTG-CG-	-TSCFKKCGG	CCDDYTGQOE	LVIKT---SH	DD-----	DD-----	
	MHV	ICCCTG-CG-	-SCCFKKCGN	CCDECGGHQD	SIVIHNISSH	ED-----	ED-----	
55	Rat CoV	ICCCTG-CG-	-SCCFKKCGN	CCDEYGGQQA	GIVIHNISSH	ED-----	ED-----	
	PHEV	ICCCTG-CG-	-TSCFKKCGG	CCDDYTGQOE	FVIKT---SH	DD-----	DD-----	
	AIBV	VFFMTGCCGC	CCGCFGIMPL	MSKCGKKSSY	YTTFDNDVVT	EQYRPKKS	VV	
	SARS	LCCMTSCCS-	-CLKGACSCG	SCCKFDEDD	EPVLKGKVLH	YT-----	YT-----	
		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		125	135	145	155	165	175	
60	EMCR 4a	FSFVLNVTK	LCFVSGKCWY	LEQSFYENRF	AAIYGGDHYV	VLGGTITFV	SFDDLYVAIR	
	229E 4a	WDFIVFNVTT	LCYAR----	-----	-----	-----	-----	

## f. Putative Orf 4a

65		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		5	15	25	35	45	55	
	EMCR 4a	MPFGLFQLT	LESTINKSVA	NLKLPPHDVT	VLRDNLKPVT	TLSTITAYLL	VSLFVTYFAL	
	229E 4a	MALG-LFTLQ	LVSANQSL	NAKVSAEVSR	QVIQDVKDGT	VTFNLLAYTL	MSLFVVYFAL	
		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		65	75	85	95	105	115	
70	EMCR 4a	FKPLTARGRV	ACFVLKLLTL	SVYVPLLVLF	GMYLDSFIIF	FLRCCFDSYM	LAIMPISNKN	
	229E 4a	FKARSHRGRA	ALIVFKILIL	FVYVPLLYWS	QAYIYATLIA	VILG-RFFH	TAWHCWLYKT	
75		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		125	135	145	155	165	175	
	EMCR 4a	FSFVLNVTK	LCFVSGKCWY	LEQSFYENRF	AAIYGGDHYV	VLGGTITFV	SFDDLYVAIR	
	229E 4a	WDFIVFNVTT	LCYAR----	-----	-----	-----	-----	

EMCR 4a  
229E 4a

.....|.....| .....|.....| .....|.....| .....|.....| .....

185            195            205            215            225

GSCEKNLQLM RKVDLYNGAV IYIFAEPPVV GIVYSSQLYE DVPSIN

-----

10

15

	..... .....	..... .....	..... .....	..... .....	..... .....	..... .....
	5	15	25	35	45	55
EMCR 4a	MPFGGLFQLT	LESTINKSVA	NLKLPPHDVT	VLRDNLKPVT	TLSTITAYLL	VSLEFVYFAL
229E 4a	MALG-LFTLQ	LVSAVNQSL	NAKVSAEVSR	QVIQDVKDGT	VTFNLLAYTL	MSLFVYFAL
229E 4b	-----	-----	-----	-----	-----	-----

		..... .....	..... .....	..... .....	..... .....	..... .....	..... .....
		65	75	85	95	105	115
20	EMCR 4a	FKPLTARGRV	ACFVLKLLTL	SVYVPLLVLV	GMYLDSFIIF	FLRCCFDSYM	LAIMPISNKN
	229E 4a	FKARSHRGRA	ALIVEKILIL	FVYVPLLWYS	QAYIYATLIA	VILLG-RFFH	TAWHCWLYKT
	229E 4b	-----	-----	-----	-----	-----	-----

25

		..... .....	..... .....	..... .....	..... .....	..... .....	..... .....
		125	135	145	155	165	175
EMCR	4a	FSFVLENVTK	LCFVSGKCWY	LEQSFYENRF	AAIYGGDHYV	VLGGETITFV	SFDDLYVAIR
229E	4a	WDFIVFNVTT	LCYAR-----				
229E	4b	-----	-----MQGKCW	FLENKALKPF	VCFYGGDQFL	YIGDRVYSYF	STNDLYVALR

30

	185	195	205	215	225
EMCR 4a	GSCEKNLQLM	RKVDLYNGAV	IYIFAEPEPVV	GIVYSSQLYE	DVPSIN
229E 4a	-----	-----	-----	-----	-----
229E 4b	GRIDKDLSL	RKVELYNGEC	VYLFCEHNAV	GIVNTDFKLE	IH...

## 40

		5	15	25	35	45	55
EMCR E	----	MFLRLI	DDNG-IVLNS	ILWLLVMIFF	F-VLAMTFFK	LIQLCFCTHY	FFSRTLYQP-
229E	----	MDLKLIV	DDHA-LVVNV	LLWCVVLLVI	L-LVCITFIK	LKLCFCFCHM	FCNRTVYGP-
PEDV	----	MLQLV	NDNG-LVVNV	ILWLFLVFFL	L-IISTTFVQ	LXNLCFCTHR	LCNSAVYTP-
45 TGEV	MTFP	PRALTVI	DDNG-MVINI	IFWELLIIL	I-LLSIALLN	IKLCMVCCN	LGRTVIIVP-
CaCoV	MTFP	PRALTVI	DDNG-MVISI	IFWELLIIL	I-LLSIALLN	IKLCMVCCN	LGRTVIIVP-
FeCoV	MTFP	PRALTII	DDHG-MVVSV	FFWLLLIIL	I-LLSIALLN	IKLCMVCCN	LGRTVIIVP-
Por Resp C	MTFP	PRALTVI	DDNG-MVISI	IFWELLIIL	I-LLSIALLN	IKLCMVCCN	LGRTVIIVP-
50 OC43	--MF	MADAYL	ADTV-WYVGQ	IIFIVAICLL	VTIIVVAFLA	TFKLCIQLCG	MCNTLVLSP-
BoCoV	--MF	MADAYF	ADTV-WYVGQ	IIFIVAICLL	VIIIVVAFLA	TFKLCIQLCG	MCNTLVLSP-
PHEV	--MF	MADAYL	ADTV-WYVGQ	IIFIVAICLL	VIIIVVAFLA	TFKLCIQLCG	MCNTLVLSP-
MHV	--MF	NLFL	TDTV-WYVGQ	IIFIVACLML	VIIIVVAFLA	TFKLCIQLCG	MCNTLVLSP-
Rat CoV	--MF	NLFL	IDTV-WYVGQ	IIFIVACLML	VIIIVVAFLA	TFKLCIQLCG	MCNTLVLSP-
AIBV	--MN	LLKSL	EEENG-SFLTA	LYIIVGFLAL	Y-LLGRALQA	FVQADACCL	FWYTVVVVPG
55 SARS	----	MYSFVS	EETGTLIVNS	VLLFLAFVVF	L-LVTLAILT	ALRLCAYCEN	IVNVSLVKKP-

	65	75	85	95	105	
EMCR E	---VYKIFL---	---	---AYQDYM---	---QIAPV-PA---	EVLNV----	--
229E	---IKNVYH---	---	---IYQSYM---	---HIDFF-PK---	RVIDF----	--
PEDV	---IGRLYR---	---	---VYKSYM---	---RIDPL-PS---	TVIDV----	--
TGEV	---AQHAYD---	---	---AYKNFM---	---RIKAYNPD---	GALLA----	--
CaCoV	---ARHAYD---	---	---AYKNFM---	---QIRAYNPD---	EALLV----	--
FeCoV	---ARHAYD---	---	---AYKTFF---	---QTKAYNPD---	EAFVL----	--
Por Resp C	---VQHAYD---	---	---AYKNFM---	---RIKAYNPD---	GALLV----	--
QC43	---SIYVFNRR GR---	---	---QFYEFYN---	---DVKPP-VL---	DVDDV----	--
BoCoV	---SIYVFNRR GR---	---	---QFYEFYN---	---DVKPP-VL---	DVDDV----	--
PHEV	---SIYVFNRR GR---	---	---QFYEFYN---	---DVKPP-VL---	DVDDV----	--
MHV	---SICVYNRR SK---	---	---QLYKYNN E---	---EVRPP-PL---	EVDIIIQTL---	--
Rat CoV	---SIYVYNRR SK---	---	---QLYKYNN E---	---EVRPP-PL---	EVDIIIQTL---	--
AIBV	AKGTAFVYKY	TYGRKLNPE	LEAVIVNEFF	KNGWNNKNPA	NFQDAQRDKL	YS
SARS	---TVVYVS---	---	---RVKNLN---	---SSEGV-PD---	LIIV-----	--

i. Putative Orf M (Matrix protein)

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|

84/07

	5	15	25	35	45	55
EMCR	-----	-----	-----	-----	-----M	SNSS-----
229E	-----	-----	-----	-----	-----M	SNDN-----
PEDV	-----	-----	-----	-----	-----M	SNGS-----
5 TGEV	-----	-----	-----MK	ILLILACVIA	CACGERYCAM	KSDTDLSCRN
CaCoV	-----	-----	-----MKK	ILFLLACVIA	CVYGERYCAM	TESS-TSCRN
FeCoV	MHMPPIRPLC	KPRHIIPTKH	FWFELNKMKY	ILLILACVIA	CVYGERYCAM	QDSG-LQCTN
PRCoV	-----	-----	-----MK	ILLILACVIA	CTCGERYCAM	KDDTGLSCRN
OC43	-----	-----	-----	-----	-----M	SSKT-----
10 PHEV	-----	-----	-----	-----	-----M	SSPT-----
BoCoV	-----	-----	-----	-----	-----M	SSVT-----
MHV	-----	-----	-----	-----	-----M	TSTTQ-----
RatSAV	-----	-----	-----	-----	-----M	SSTTP-----
AIBV	-----	-----	-----	-----	-----M	PNETN-----
15 SARS	-----	-----	-----	-----	-----M	ADNG-----
	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	65	75	85	95	105	115
20 EMCR	-----V	PLSEVYVHLR	NWNFSWNLLI	TVFIVVLQYG	HYKYSRLLYG	LKMSVLWCLW
229E	-----C	-TGDIVTHLK	NWNFGWNVIL	TIFIVILQFG	HYKYSRLFYG	LKMLVLWLLW
PEDV	-----I	PVDEVIEHLR	NWNFTWNIIL	TILLVVLQYG	HYKYSVFLYG	VKMAILWILW
TGEV	STASDCESCF	NGGDLIWHLA	NWNFSWSIIL	IVFITVLQYG	RPQFSWFVYG	IKMLIMWLLW
CaCoV	STAGNCASCF	ETGDLIWHLA	NWNFSWSVIL	IFFITVLQYG	RPQFSWFVCG	IKMLIMWLLW
FeCoV	GTNSRCQTCF	ERGDLIWHLA	NWNFSWSVIL	IVFITVLQYG	RPQFSWLVYG	IKMLIMWLLW
25 PRCov	GTASDCESCF	NRGDLIWHLA	NWNFSWSIIL	IFFITVLQYG	RPQFSWFVYG	IKMLIMWLLW
OC43	--TPAPVYIW	TADEAIKFLK	EWNFSLGIIIL	LFITIIILQFG	YTSRSMFVYV	IKMIILWLMW
PHEV	--TPVPVISW	TADEAIKFLK	EWNFSLGIIIV	LFITIIILQFG	YTSRSMFVYV	IKMVILWLMW
BoCoV	--TPAPVYTW	TADEAIKFLK	EWNFSLGIIIL	LFITIIILQFG	YTSRSMFVYV	IKMIILWLMW
MHV	--APQPXYQW	TADEAIRFLK	EWNFSLGIIIL	LFITIIILQFG	YTSRSMFVYV	VKMILLWLMW
30 RatSAV	--APQTVYQW	TADVAVRFLK	EWNFSLGIIIL	LFITIIILQFG	YTSRSMFVYV	VKMILLWLMW
AIBV	-----CTL	DFAQSVQLEK	EYNLFITAFI	LFITIIILQYG	YATRSKVITY	LKMIIVLWCFW
SARS	-----TI	TVEELKQLE	QWNLVIGFLF	LAWIMLLQFA	YSNRNRFYI	IKLVFLWLLW
	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	125	135	145	155	165	175
35 EMCR	PLVLALSIFD	CFVNFENV-D	VFFGFSILMS	IITLCLWVMY	FVNSFRLWRR	VKTFAENFPE
229E	PLVLALSIFD	TWANWDSN-W	AFVAFSFFMA	VSTLVMWVMY	FANSFRLFR	ARTFAWNPE
PEDV	PLVLALSIFD	AWASEFQVN-W	VFFAFSILMA	CITLMLWIMY	FVNSIRLWRR	THSWWSFNPE
40 TGEV	PVVLALTIFN	AYSEYQVSRY	VMFGFSIAGA	IVTFVLWIMY	FVRSIQLYRR	TKSWSFNPE
CaCoV	PIVLALTIFN	AYLEYRVSRY	VMFGFSVAGA	TVTFILWIMY	FVRSIQLYRR	TKSWSFNPE
FeCoV	PIVLALTIFN	AYSEYQVSRY	VMFGFSVAGA	VVTFALWMMY	FVRSVQLYRR	TKSWSFNPE
PRCoV	PIVLALTIFN	AYSEYQVSRY	VMFGFSIAGA	IVTFVLWIMY	FVRSIQLYRR	TKSWSFNPE
OC43	PLTIILTIFN	--CVYALN-N	VYLGSIVFT	IVAIMMWIVY	FVNSIRLFIR	TGSWSFNPE
PHEV	PLTIILTIFN	--CVYALN-N	VYLGSIVFT	IVAIMMWIVY	FVNSIRLFIR	TGSWSFNPE
45 BoCoV	PLTIILTIFN	--CVYALN-N	VYLGSIVFT	IVAIMMWIVY	FVNSIRLFIR	TGSWSFNPE
MHV	PLTIVLCIFN	--CVYALN-N	VYLGSIVFT	IVSIVMWIMY	FVNSIRLFIR	TGSWSFNPE
RatSAV	PLTIVLCIFN	--CVYALN-N	VYLGSIVFT	IVSIVMWIMY	FVNSIRLFIR	TGSWSFNPE
AIBV	PLNIAVGVIS	--CTYPPN-T	GGLVAAIILT	VFACLSFVG	WQISIRLFKR	CRSWSFNPE
50 SARS	PVTLACFVLA	--AVYRIN-W	VTGGIAIAMA	CIVGLMWLSY	FVASFRLFR	TRSMWSFNPE
	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	185	195	205	215	225	235
55 EMCR	TNAIISLQVY	-GHNYYLPVM	AAPTGVTLTL	LSGVLLVDGH	KIATRVQVGO	LPKYVIVATP
229E	VNAITVTVL	-GQTYOPIQ	QAPTGVTLTL	LSGVLYVDGH	RLASGVQVHN	LPEYMTVAVP
PEDV	TDALLTSVM	-GROVCPEVL	GAPTGVTLTL	LSGTLVEGY	KVATGVQVSO	LPNFVTVAKA
TGEV	TKAILCVSAL	-GRSYVLEPL	GVPTGVTLTL	LSGNLYAEGF	KIAGGMNIDN	LPKYVMVALP
CaCoV	TSAILCVSAL	-GRSYVLEPL	GVPTGVTLTL	LSGNLYAEGF	KIAGGMNIDN	LPKYVMVALP
FeCoV	TNAILCVNAL	-GRSYVLEPL	GVPTGVTLTL	LSGNLYAEGF	KIAGGLTIEH	LPKYVMIATP
PRCoV	TNAILCVSAL	-GRSYVLEPL	GVPTGVTLTL	LSGNLYAEGF	KIAGGMNIDN	LPKYVMVALP
60 OC43	TNNLMCIDMK	-GRMYVRPII	EDYHTLTATI	IRGHLYIQGI	KLGTGYSLSL	LPAYMTVAK-
PHEV	TNNLMCIDMK	-GRMYVRPII	EDYHTLTATI	IRGHLYIQGI	KLGTGYSLSL	LPAYMTVAK-
BoCoV	TNNLMCIDMK	-GRMYVRPII	EDYHTLTATI	IRGHLYMQGI	KLGTGYSLSL	LPAYMTVAK-
MHV	TNNLMCIDMK	-GTIVYVRPII	EDYHTLTATI	IRGHLYMQGV	KLGTGYSLSL	LPAYMTVAK-
RatSAV	TNNLMCIDVK	-GTIVYVRPII	EDYHTLTATI	IRGHLYMQGV	KLGTGYSLSL	LPAYMTVAK-
65 AIBV	SNAVGSILIT	NGQQCNFAIE	SVPMLVSPII	KNGVLYCEGQ	WLAK-CEPDH	LPKIDFVCTP
SARS	TNILLNVPLR	-GTIVTRPLM	ESELVIGAVI	IRGHLRMAGH	PLGR-CDIKD	LPKEITVAT-
	.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
	245	255	265	275	285	
70 EMCR	STTIVCDRVG	RSVNETSQTG	WAFYVRKAGH	DFSQVAVQEG	VLSEREKLLH	LI
229E	STTIYSRVG	RSVNSQNSTG	WVFYVRVKHG	DFSQVSSPMS	NMTENERLLH	FF
PEDV	TTTIVYGRVG	RSVNASSGTG	WAFYVRSKHG	DYSAVSNPSA	VLTDSKVLH	LV
TGEV	SRTIVYTLVG	KKLKASSATG	WAYVYVSKAG	DYSTEAR-TD	NLSEHEKLLH	MV
CaCoV	VRTIVYTLVG	KKLKASSATG	WAYVYVSKAG	DYSTEAR-TD	NLSEHEKLLH	MV
75 FeCoV	SRTIVYTLVG	KKLKASSATG	WAYVYVSKAG	DYSTEAR-TD	NLSEHEKLLH	MV
PRCoV	SRTIVYTLVG	KKLKASSATG	WAYVYVSKAG	DYSTEAR-TD	NLSEHEKLLH	MV
OC43	VTHLCTYKRG	FLDRISDTSG	FAVYVSKVKG	NYRLPSTQKG	SGMDTALLRN	NI
PHEV	VTHLCTYKRG	FLDRIGDTSG	FAVYVSKVKG	NYRLPSTHKG	SGMDTALLRN	NI
BoCoV	VSHLLTYKRG	FLDKIGDTSG	FAVYVSKVKG	NYRLPSTQKG	SGMDTALLRN	NI
80 MHV	VSHLCTYKRA	FLDKVDGVSG	FAVYVSKVKG	NYRLPSN-KP	SGMDTALLR-	-I

RatSAV	VSHLCTYKRA	FLDKVDGVSG	FAVYVKSXVG	NYRLPSN-KP	SGADTALLR-	-I
AIBV	DRRNIRMVQ	KYTGDSQGNK	KRFATFVYAK	QSVDTGELES	VATGSSSLYT	--
SARS	SRTLSTYYKLG	ASQRVGTDSG	FAAYNRYRIG	NYKLNTDHAG	SNDNIALLVQ	--

5

## 10 j, Putative Orf N (Nucleoprotein)

	5	15	25	35	45	55
EMCR	-----	--MAS-----	---VN---W	ADDR-----	--AARKKF--	-----
229E	-----	--MAT-----	---VK---W	ADASEPQ---	--RGRQGR--	-----
PEDV	-----	-----	---VS---F	QDRG-----	--RKR-----	-----
TGEV	-----	--MANQGQR-	---VS---W	GDESTKT---	--RGRSNSRG	RKNNN-----
FeCoV	-----	--MATQGQR-	---VN---W	GDEPSKR---	--RGRSNSRG	RKNNN-----
PRCoV	-----	--MANQGQR-	---VS---W	GDESTKI---	--RGRSNSRG	RKINN-----
CaCoV	-----	--MASQGQR-	---VS---W	GDESTKR---	--RGRSNSRG	RKNNN-----
RSDACoV	MSFVPGQENA	GSRSSSGNRA	NGILKKTW	ADQTERGQNN	GNGRRNQPK	QTATTQ-PNT
MHV	MSFVPGQENA	GSRSSSGNRA	NGILKKTW	ADQTERG---	--NRGRNHPK	QTATTQ-PNA
PHEV	MSFTPGKQSS	-SRASSGNRS	NGILK---	ADQSDQSRNV	QTRGRRVQSK	QTATSQQPSG
OC43	MSFTPGKQSS	-SRASSGNRS	NGILK---	ADQSDQFRNV	QTRGRRAPK	QTATSQQPSG
BoCoV	MSFTPGKQSS	-SRASFNNRS	NGILK---	ADQSDQSRNV	QTRGRRAPK	QTATSQQPSG
SARS	-----	--MSDNGFQS	NQRSAPRITF	GGPTDSTDNN	QNGRRNGARP	KQRRPQ---
AIBV	-----	--MASG-----	---K---A	AGKTDAPAPV	IKLGGPKPPK	VGSS-----

	65	75	85	95	105	115
EMCR	-----	PPPSFY	MPLLVSSDKA	PYRVIPRNLV	PIGKGNK-DE	QIGYWNVQER
229E	-----	IPYSLY	SPLLVDSE-Q	PWKVIPRNLV	PINKKDK-NK	LIGYWNVQKR
PEDV	-----	VPLSLY	APLRVTNDKP	LSKVLANNV	PTNKGNK-DQ	QIGYWNVQIR
TGEV	-----	IPLSFF	NPITLQGGSK	FWNLCPRDFV	PKGIGNR-DQ	QIGYWNVQTR
FeCoV	-----	IPLSFF	NPITLQGGSK	FWNLCPRDLV	PKGIGNR-DQ	QIGYWNVQIR
PRCoV	-----	IPLSFF	NPITLQGGSK	FWNLCPRDFV	PKGIGNR-DQ	QIGYWNVQTR
CaCoV	-----	IPLSFF	NPITLQGGSK	FWNLCPRDFV	PKGIGNR-DQ	QIGYWNVQTR
RSDACoV	GSVVPYHSWF	SGITQFQKKG	EFQFAGQGV	PIANGIPPSE	QKGYWYRHNH	RSFKTPDQGO
MHV	GSVVPYHSWF	SGITQFQKKG	EFQFAGQGV	PIANGIPPSE	QKGYWYRHNH	RSFKTPDQGO
PHEV	GTVPYYSWF	SGITQFQKKG	EFQFAGQGV	PIAPGVASTE	AKGYWYRHNH	RSFKTADGNQ
OC43	GTVPYYSWF	SGITQFQKKG	EFQFAGQGV	PIAPGVASTE	AKGYWYRHNH	RSFKTADGNQ
BoCoV	GNVVPYYSWF	SGITQFQKKG	EFQFAGQGV	PIAPGVASTE	AKGYWYRHNH	RSFKTADGNQ
SARS	GLPNNTASWF	TALTOHKG-E	ELRFPGRQGV	PINTNSGPD	QIGYWRATR	R-VRGDQGM
AIBV	-----	GNASWF	QAIAKAKLNT	PPPKEFGSGV	PDNENIKPSQ	QHGYYWRQAR

	125	135	145	155	165	175
EMCR	VDLPEKVFHY	YLGTGPHKDL	KFRQRSDGVV	WVAKEGAKTV	NTSLGNRK--	RNOKPLEPKF
229E	VDLSPKLFHY	YLGTGPHKDA	KFRERVEGVV	WVAVDGAKTE	PTGYGVRR--	KNSEPEIPHF
PEDV	IEQPSNWHFY	YLGTGPHGDL	RYRTRTGVF	WVAKEGAKTE	PTNLGVVK--	ASEKPIIPKF
TGEV	KELPERWFFY	YLGTGPHADA	KFKDKLDGVV	WVAKDGMANK	PTTLGSRG--	ANNESKALKF
FeCoV	KELPERWFFY	YLGTGPHADA	KFKDKLDGVV	WVAKDGMANK	PTTLGSRG--	ANNESKALKF
PRCoV	KELPERWFFY	YLGTGPHADA	KFKDKLDGVV	WVAKDGMANK	PTTLGSRG--	ANNESKALKF
CaCoV	KNLPEKWFY	YLGTGPHADA	KFKDKLDGVV	WVAKDGMANK	PTTLGSRG--	ANNESKALKF
RSDACoV	KQLLPRWYFY	YLGTGPHAGA	SFGDSIEGVF	WVASQADTN	TSADIVERDP	SSHEAIPTRF
MHV	KQLLPRWYFY	YLGTGPHAGA	EYGDIDGVV	WVASQADTN	TSADIVERDP	SSHEAIPTRF
PHEV	RQLLPRWYFY	YLGTGPHAKD	QYGTIDGVV	WVASQADTN	TPADIVDRDP	SSDEAIPTRF
OC43	RQLLPRWYFY	YLGTGPHAKD	QYGTIDGVV	WVASQADTN	TPADIVDRDP	SSDEAIPTRF
BoCoV	RQLLPRWYFY	YLGTGPHAKD	QYGTIDGVV	WVASQADTN	TPADIVDRDP	SSDEAIPTRF
SARS	KELSPRWYFY	YLGTGPEASL	FYGANKGIV	WVATEGALNT	PKDHIGTRNP	NNNAATVLQL
AIBV	KPVDPDAWYFY	YTGTGPAADL	NWGDQDQGV	WVAAGADTK	SRSNQGTDRP	DKFDQYPLRF

	185	195	205	215	225	235
EMCR	SIALPPELSV	VEFEDRSNNS	SRASSRSSTR	-----	-----	-----
229E	NQKLPGVTV	VEEPD----	SRAPSRQSR	-----	-----	-----
PEDV	SQQLPSVVEI	VEPNTTP--A	SRANSRSRSP	GNGNNRSP	SNNRGNNQSR	GNSQNRGNQ
TGEV	DGKVPGEFQL	EVNQS----	RDNSRSRSQ	-----	-----	-----
FeCoV	DGKIPPPQQL	EVNRS----	RNNSRSRSQ	-----	-----	-----
PRCoV	DGKVPGEFQL	EVNQS----	RDNSRSRSQ	-----	-----	-----
CaCoV	DGKVPGEFQL	EVNQS----	RDNSRSRSQ	-----	-----	-----
RSDACoV	DVGVPEFHL	EVNQL----	RDNSRSRSQ	-----	-----	-----
MHV	APGTVLPQGF	YVEGS----	GRSAPASRS	-----	-----	-----
PHEV	APGTVLPQGF	YVEGS----	GRSAPASRS	-----	-----	-----
OC43	PPGTVLPQGY	YIEGS----	GRSAPNSRST	-----	-----	-----
BoCoV	PPGTVLPQGY	YIEGS----	GRSAPNSRST	-----	-----	-----
SARS	PPGTVLPQGY	YIEGS----	GRSAPNSRST	-----	-----	-----
AIBV	PQGTTLKPGF	YAEGS----	RGGSQASSR	-----	-----	-----
	SDGGPDGNFR	WDFIP----	LN-RGRSG	-----	-----	-----

	245	255	265	275	285	295
--	-----	-----	-----	-----	-----	-----

86/87

5	EMCR	SRSTSRQ---	-----QSR-	TRSDSNQS--	-----S-SDL	VAAVTLALKN	LGFDN--QSK
	229E	GRGESKP---	-----QSRN	PSSDRNHN--	-----SQDDI	MKAVAAALKS	LGFDKP-QEK
	PEDV	GRGASQNRGG	NNNNNNKSRN	QSNNRNQSND	RGGVTSRDDL	VAAVKDALKS	LGIGEN-PDR
	TGEV	SRSRSRNR--	-----SQSRG	RQFFNNKK--	-----DDSV	EQAVLAALKK	LGVDTE-KQQ
	FeCoV	SRSVSRNR--	-----SQSRG	RHHSNNQ--	-----NNNV	EDTIVAVLEK	LGV-TD-KQ-
	PRCoV	SRSRSRNR--	-----SQSRG	RQOSNNKK--	-----DDSV	EQAVLAALKK	LGVYTE-KQQ
	CaCoV	SRSQSRNR--	-----SQSRG	RQLSNKKK--	-----DDNV	EQAVLAALKK	LGVDTE-KQQ
	RSDACoV	SRSQSRGP--	-----NNRA	RSSSNQORQ--	-----PASTV	KPDMAEEIAA	LVLAN---LG
10	MHV	SRSQSRGP--	-----NNRA	RSSSNQORQ--	-----PASAV	KPDMAEEIAA	LVLAK---LG
	PHEV	SRAPNRAPS--	-----AGSRS	RANSGNRT--	-----STPGV	TPDMADQIAS	LVLAK---LG
	OC43	SRTSSRASS--	-----AGSRS	RANSGNRT--	-----PTSGV	TPDMADQIAS	LVLAK---LG
	BoCoV	SRASSRASS--	-----AGSRS	RANSGNRT--	-----PTSGV	TPDMADQIAS	LVLAK---LG
	SARS	SSRSRSGN--	-----SRNST	PGSSRGNS--	-----PARMA	SGGETALAL	LLDLRLNQL
	AIBV	-RSTAASS--	-----AAASRA	PSREGSRG--	-----RRSDS	GDDLIARAAK	IIQDQ-----
15		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		305	315	325	335	345	355
20	EMCR	SPSSSGTSTP	K-----K-	-----PNKPLSQ	PRADKPS---	-QLKKPRWKR	VPTR--EENV
	229E	DKKSAKTGTP	KPSRNQSPAS	SQTSAKSLAR	SQSSETEKQK	HEMQKPRWKR	QPNDDVTSNV
	PEDV	HKQQQKPKQE	K-SDN---SG	KNTPKKNKSR	ATSKERD---	-LKDIPEWRR	IPKG--ENSV
	TGEV	QRSRSKSKER	S-----S	-----NSKTR	DTPPKNE---	---NKHTWKR	TAGK---GDV
	FeCoV	-RSRSKPRER	S-----S	-----DSKPR	DTPPKNA---	---NKHTWKK	TAGK---GDV
	PRCoV	QRSRSKSKER	S-----S	-----NSKTR	DTPPKNE---	---NKHTWKR	TAGK---GDV
	CaCoV	-RSRSKSKER	S-----S	-----SSKTR	DTPPKNE---	---NKHTWKR	TAGK---GDV
25	RSDACoV	-KDAGQPKQV	T-----T	-----KQSAK	EVROKIL---	---NKPRQKR	TPNK--QCPV
	MHV	-KDAGQPKQV	T-----T	-----KQSAK	EVROKIL---	---TKPRQKR	TPNK--QCPV
	PHEV	-KDATKPKQV	T-----T	-----KQSAK	EVROKIL---	---NKPRQKR	SPNK--QCTV
	OC43	-KDATKPKQV	T-----T	-----KHTAK	EVROKIL---	---NKPRQKR	SPNK--QCTV
	BoCoV	-KDATKPKQV	T-----T	-----KQSAK	EVROKIL---	---NKPRQKR	SPNK--QCTV
30	SARS	SKVSGKGQQQ	Q-----Q	-----GQTVTK	KSAAEAS---	---KKPRQKR	TATK--QYNV
	AIBV	---QKKSRI	T-----T	-----KAKAD	EMAHRRY---	---CKRT	IPPN--YRV
35		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		365	375	385	395	405	415
	EMCR	IQCFGPRDFN	H---NMGDS	LVQNGVDAG	FPQLAELIPN	QAALFFDSEV	STDEVG----
	229E	TQCFGPRDL	H---NFGSAG	VVANGVAKAG	YPQFAELVPS	TAAMLFDSHI	VSKEG----
	PEDV	AACFGPRGGF	K---NFGDAE	FVEKGVDSAG	YAQIASLAPN	VAALLFGGNV	AVRELA----
	TGEV	TRFYGARSSS	A---NFGD	LVANGSSAKH	YPQLAECVPS	VSSILFGSYW	TSKEDG----
	FeCoV	TRFYGARSSS	A---NFGD	LVANGSSAKH	YPQLAECVPS	VSSILFGSYW	SAEEAG----
40	PRCoV	TRFYGARSSS	A---NFGD	LVANGSSAKH	YPQLAECVPS	VSSILFGSYW	TSKEDG----
	CaCoV	TKFYGARSSS	A---NFGD	LVANGSSAKH	YPQLAECVPS	VSSILFGSHW	TAKEDG----
	RSDACoV	QQCFGKRGPN	Q---NFGGPE	MLKLGTSDPQ	FPILAEALPT	PGAFFFGSKL	ELVKKN--SG
	MHV	QQCFGKRGPN	Q---NFGGSE	MLKLGTSDPQ	FPILAEALPT	PSAFFFGSKL	ELVKKN--SG
	PHEV	QQCFGKRGPN	Q---NFGGSE	MLKLGTSDPQ	FPILAEALPT	AGAFFFGSRL	ELAKVQNLSS
45	OC43	QQCFGKRGPN	Q---NFGGSE	MLKLGTSDPQ	FPILAEALPT	AGAFFFGSRL	ELAKVQNLSS
	BoCoV	QQCFGKRGPN	Q---NFGGSE	MLKLGTSDPQ	FPILAEALPT	AGAFFFGSRL	ELAKVQNLSS
	SARS	TQAFGRRGPE	QTQGNFGDQD	LIRQGTQYKH	WPQIAQFAPS	ASAFFFGMSRI	GMEVTP----
	AIBV	DQVFGPRTKG	K-EGNFGDDK	MNEEGIKDGR	VTAMLNLVPS	SHACLFGRSV	TPKLQL----
50		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		425	435	445	455	465	475
	EMCR	-----DNV	QITYT---Y	KMLVAKDNKN	LPKFIEQISA	FTKPS-----	SIKEMSQSS
	229E	-----NTV	VLTFE---T	RVTVPKDHPP	LKGFLEELNA	FTR-----	EMQOHPLLNP
	PEDV	-----DSY	EITYN---Y	KMTVPKSDPN	VELLVSVQDA	FKTGN-----	AKLQKKEKK
55	TGEV	-----DQI	EVTFT---H	KYHLPKDDPK	TGQFLQIQINA	YARPS-----	EVAKEQRRK
	FeCoV	-----DQV	KVTLT---H	TYLPLKDDAK	TSQFLEQIDA	YKRPS-----	EVAKDQRRK
	PRCoV	-----DQI	EVTFT---H	KYHLPKDHPP	TEQFLQIQINA	YASPS-----	ELAKEQRRK
	CaCoV	-----DQI	EVTFT---H	KYHLPKDDPK	TGQFLQIQINA	YARPS-----	EVAKEQRRK
60	RSDACoV	GVDEPTKDVY	ELQYSGAVRF	DSTLPGFETI	MKVLNENLNA	YQNOA-----	GGADVVSPPK
	MHV	GADEPTKDVY	ELQYSGAIRF	DSTLPGFETI	MKVLNENLNA	YQDQA-----	GSDVLVSPKP
	PHEV	NPDEPQKDVY	ELRYNGAIRF	DSTLSGFETI	MKVLNENLNA	YQHQA-----	DGMNISPSPK
	OC43	NPDEPQKDVY	ELRYNGAIRF	DSTLSGFETI	MKVLNENLNA	YQQQ-----	DGMNISPSPK
	BoCoV	NLDEPQKDVY	ELRYNGAIRF	DSTLSGFETI	MKVLNENLNA	YQQQ-----	DGMNISPSPK
65	SARS	-----SGT	WLTGHGAIKL	DDKDPQFKDN	VILLNKHIDA	YKTFP-----	PTEPKKDKKK
	AIBV	-----DGL	HLRFETTVV	PCDDPQFDNY	VKICDQCVDG	VGTRPKDDEP	KPKSRSSSRP
70		.... ....	.... ....	.... ....	.... ....	.... ....	.... ....
		485	495	505	515	525	535
	EMCR	HVAQNTVLN-	-----	-ASIPES---	-KPLADDDSA	IIIEIVNEVLH	-----
	229E	SALEFNPSQ-	-----	-TSPATA---	-EPVRDEVSI	ETDIIDEVN-	-----
	PEDV	NKRETTLQOH	EEAIYDDVGA	PSDVTHANLE	WDTAVDGGDT	AVEIINEIFD	TGN-----
	TGEV	SRSKSAERS-	-----	EQDVVPDALI	ENYTDVFDDT	QVEIIDEVTN	-----
	FeCoV	SRSKSADK--	-----	KPELSVTLV	EAYTDVFDDT	QVEMIDEVTN	-----
	PRCoV	SRSKSAERS-	-----	EQEVVPDSLI	ENYTDVFDDT	QVEMIDEVTN	-----
75	CaCoV	ARSKSVERV-	-----	EQEVVPDALI	ENYTDVFDDT	QVEIIDEVTN	-----
	RSDACoV	QRKRGTKQT-	-AQKEELDSI	SVAKPKSAVQ	RNVSRELTPE	DRSLLAQILD	DGVVPDGLDD
	MHV	PRRGRRQAQ-	-EKKDEVNDV	SVAKPKSLVQ	RNVSRELTPE	DRSLLAQILD	DGVVPDGLDD
	PHEV	QRQRGQKN--	--GQVENDNV	SVAAPKSRVQ	QNKSRRELTAE	DISLLKKMDE	P-----YTED
	OC43	QRQRGHKN--	--GQVENDNI	SVAAPKSRVQ	QNKSRRELTAE	DISLLKKMDE	P-----YTED
80	BoCoV	QRQRGQKN--	--GQVENDNI	SVAAPKSRVQ	QNKSRRELTAE	DISLLKKMDE	P-----YTED

87/87

	SARS	KTDEAQPLP- -----	QRQKKQPTVT	LLPAADMDDF	SRQLQNSMSG	ASADSTQA--
	AIBV	ATRGNSPAPR	QQRPKKEKKL	KKQDDEADKA	LTSDEERNNA	QLEFYDEPKV INWGDAAALGE
5		....				
	EMCR	-----				
	229E	-----				
	PEDV	-----				
	TGEV	-----				
10	FeCoV	-----				
	PRCoV	-----				
	CaCoV	-----				
	RSDACoV	-SNV				
	MHV	DSNV				
15	PHEV	TSEI				
	OC43	TSEI				
	BoCoV	TSEI				
	SARS	-----				
20	AIBV	NEL-				

25 k. 5'untranslated region (genomic sequence)

30	EMCR5'UTR	5	15	25	35	45	55
	229E5'UTR	ACTTAAGTAC	CTTATCTATC	TACAGATAGA	AAAGTTGCTT	-TTTAGACTT	TGTGTCTACT
35	EMCR5'UTR	65	75	85	95	105	115
	229E5'UTR	CCTCTCAACT	AAACGAAATT	TTT-CTAGTG	CTGTCATTG	TTATG--GCA	GTCCTAGTGT
		TTTCTCAACT	AAACGAAATT	TTTGCTATGG	CCGGCATCTT	TGATGCTGGA	GTCGTAGTGT
40	EMCR5'UTR	125	135	145	155	165	175
	229E5'UTR	AATTGAAATT	TCGTCAAGTT	TGTAA-ACTG	GTTAGGCAAG	TGTTGTATT	TCTGTGTTTA
		AATTGAAATT	TCATTTGGGT	TGCAACAGTT	TGGAAGCAAG	TGCTGTGTGT	CCTA-GTCTA
45	EMCR5'UTR	185	195	205	215	225	235
	229E5'UTR	AGCAC TGGTG	GTTCTGTC-C	ACTAGTGCAC	AC-ATTGATA	CTTAAGT-GG	TGTTCTGTCA
		AGGGTTTCGT	GTTCCGTCAC	GAGATTCCAT	TCTACAAACG	CCTTACTCGA	GGTCCGTCT
50	EMCR5'UTR	245	255	265	275	285	
	229E5'UTR	CTGCTTATTG	TGGAAGCAAC	GTTCTGTCTT	TGTGGAAACC	AATAACTGCT	AACC
		CGTGTTTGTG	TGGAAGCAAA	GTTCTGTCTT	TGTGGAAACC	AGTAACTGTT	CCTA

55

# Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/NL04/000805

International filing date: 18 November 2004 (18.11.2004)

Document type: Certified copy of priority document

Document details: Country/Office: EP  
Number: 03078772.5  
Filing date: 01 December 2003 (01.12.2003)

Date of receipt at the International Bureau: 31 January 2005 (31.01.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)



World Intellectual Property Organization (WIPO) - Geneva, Switzerland  
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse